

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Thackeray, Russell (2017) A new conceptual framework for the evaluation of L&D programmes.
DProf thesis, Middlesex University. [Thesis]

Final accepted version (with author's formatting)

This version is available at: <https://eprints.mdx.ac.uk/22694/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

A New Conceptual Framework for the Evaluation of L&D Programmes

A project submitted to Middlesex University to meet part
of the requirements for the degree of
Doctor of Professional Studies
(Leadership Development and Executive Coaching)

Russell Thackeray

Student Number **M00337728**

Institute for Work Based Learning
Middlesex University
December 2016

Disclaimer

The views expressed in this document are mine and are not necessarily the views of my supervisory team, examiners or Middlesex University.

Copyright

Attention is drawn to the fact that copyright of this document rests with the author. This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that copyright rests with the author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author. This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Abstract

The evaluation of learning and development programmes is one of the most important processes for the L&D function as it determines functional credibility by the creation of value from the L&D budget.

The Kirkpatrick (1954) framework has primacy in the world of evaluation along with a small number of other frameworks, e.g. Phillips (1999), Holton (2000) Brinkerhoff (2003). However, despite its acknowledged ineffectiveness, efforts to replace the framework have been unsuccessful, becoming the prevailing paradigm for evaluation which could be an increasing risk for L&D itself in proving legitimacy and credibility.

A review of wider literature reveals new opportunities to synthesise different theoretical positions to build a new framework that could add value to practice, particularly by using Decision Theory, Hubbard (2007), Forecasting, Armstrong (2001) and 'Intention' Anscombe (1959). This study adds to the literature by highlighting opportunities from that synthesis for further research and for practice.

The researcher is a specialist and experienced evaluator of L&D programmes and functions with access to a wide range of practitioners and organisations willing to help with research.

This research investigates attitudes and current practice in evaluation and, using new technologies and the synthesis of ideas and methods, to posit a new evaluation framework. This framework builds an evaluation strategy using forecasting methods with the Central Limit Theorem as the key drivers in its evaluation 'mix' to generate highly accurate outputs. The framework has a subset of Quadrants that generate the evaluation outcomes and one specific Quadrant is the subject of this research, assessed using a Case Study approach to be shown to have potential impact for L&D.

The findings from the research show that this new framework can deliver evaluation outputs with targeted levels of accuracy for a fraction of the cost, time and resource required by the traditional summative methods, currently used as part of the existing evaluation paradigm.

Whether this approach can rival the prevailing paradigm will be for practitioners to decide and is outside the scope of this research but it is suggested that it could offer a real choice for L&D evaluators.

Word Count: 344 Words

Acknowledgements

The process of conducting research is rarely down to the efforts of one person and, although the research seemed lonely and removed from real life from time to time, this is a welcome opportunity for me to recognise the efforts and contribution of others without whom this project would simply not have been possible.

Firstly, having my research supervisor, Annette Fillery-Travis on my side was a blessing. I have so enjoyed our interactions and she has helped by successfully challenging my thinking and developed my approach by guiding, coaching and cajoling; using a range of techniques from 'temptation' to deployment of 'the steely gaze', to encourage my reflection to develop my position both as a learner as well as a coach.

To the many practitioners who helped in this research, I owe a tremendous debt. From the initial sample, through to those amazing six people who worked with me on the Case Study, thank you for trusting me enough to experiment in your organisations with my new framework. I realise you all took a risk and I hope I can thank you all more publicly in due course.

Colleagues at work deserve thanks and have been a tremendous support in helping me have the time and space to disappear for periods to 'do the doctorate', and their enthusiasm and support made life much easier.

I also had the contribution of Valerie Anderson and David Rock as consultants provided by the university. Both added tremendous value from their inputs – I wish more contact had been possible. Also Lorraine Boul, as my professional 'critical friend', deserves thanks for frequent telephone calls, ideas, challenges and sharing wisdom from time to time that kept me engaged in the process.

The adage that 'behind a successful man sits a powerful woman' is completely true in my experience. My wife has been the single most

important person in this process. When I became distracted, she was able to cajole me into returning to the computer screen. When I was becoming frustrated with the whole process and threatening to just give up, she was able to re-engage my curiosity and enthusiasm for research. When struggling with concepts, she engaged me with her own thoughts and ideas and listened to endless renditions of stylistic changes and readings of versions of extremely tedious text. She has also cheerfully listened to my ideas, frustrations, cursing, and provided tea, food, inspiration and support, to ensure I could deliver this document. Without her, none of this would have been possible or achieved.

Glossary

The '***Bernoulli approach***' is the concept developed by Jacob Bernoulli and later described as the 'Law of Large Numbers' where substantial amounts of data always reflect a regression to the mean when analysed.

The 2002 ***Brinkerhoff Success Case Method*** is a purposive (and deliberately small) sample of a group of delegates from which can be extracted both evidence of impact as well as opinion about the effectiveness of learning.

The ***Chartered Institute of Personnel and Development*** (CIPD) is the professional body for Human Resources, L&D and Organisational Development practitioners. They are responsible for accreditation of membership grades, professional qualifications and professional development.

Decision Theory is a term often applied to a wide range of concepts and approaches. For the purposes of this research paper, I have followed the frameworks proposed by Steele (2015) and Hubbard (2007. 2014). According to Steele et al. (2015), Decision Theory is usually broken into two drivers - one that concerns the ability to make the 'best' decisions (normative), and the other that explains how people actually 'make' decisions (descriptive). It is the normative branch that is referred to within in this document. The basic principle according to Hubbard (2007) is to begin with a decision and then choose an appropriate process to determine how to make that decision. Whether that process simply serves to create a probability for the decision to be correct or to reduce uncertainty about the outcome, the starting place is always to define the problem in hand. Hubbard (2014) also states that, as part of any decision, a process should be used to calculate that the value of the decision must be worth the cost of the decision process.

Evaluation

The Oxford English dictionary defines evaluation as

The making of a judgement about the amount, number, or value of something; assessment.

Within L&D, evaluation often means ensuring that some output of learning can be measured and assessed after the learning intervention and to be able to ascribe a value to the intervention itself.

The evaluation paradigm was created from the use of the Kirkpatrick (1954) four levels that aims to evaluate behaviours, reaction, how much learning has been retained and results post course. The levels aim to give a structure in order to collect the data in order to analyse the findings to determine that value.

This report is the investigation into the advances and processes involved in evaluation and how the use of evaluative techniques could change in the future.

The '**Jack Phillips Method**' from the ROI Academy is often referred to as the 'Gold Standard' summative evaluation method, as it creates a robust calculation of financial Return on Investment. It is particularly rigorous in allocating costs to the training process as well as allowing for regular data collection from a range of learning stakeholders to generate a value from learning outcomes and impacts.

Learning and Development

This term refers to the department within an organisation responsible for the direct or indirect delivery of a range of learning interventions. Originally known as the Training and Development function, it has changed its name following the rebranding of HR by Ulrich (1997) to reflect the fact that is involved in the supply of interventions other than traditional training. These can include a wider range of technology solutions as well as coaching and blended learning and wider conferences and events. The name change also reflected the removal of Organisation Development functions as those processes have often become part of the

L&D function, for example, Learning Needs Analysis, Learning Design and Learning Evaluation

A ***Learning Management System*** ('LMS') is the core software used to administrate the main processes of the L&D team. It can be used to schedule and book people onto courses and handle the distribution of communication and learning inputs. The LMS often operates the post course administration of data collection to assist with evaluation.

Training

Within an organisational, according to Kraiger (1993) training refers to the particular process involved in advancing the skills, knowledge, or attitudes of an employee. Most training satisfies the requirement that the training is needed to ensure that employees can operate organisational processes against a range of factors including competence, legal, productivity, or quality requirements. Some training is based solely around specific job requirements and is provided by the organisation whilst other training can be part of a professional qualification route and supplied by a third party (for example Legal Qualifications). Increasingly, there has been a move to train people in 'softer skills' to develop motivation and morale driven by the growth of advances in thinking around HRM. Traditional training takes place in a group setting away from the job or by a supervisor 'on the job' or by using a technology solution depending on the level of the learner, complexity of the knowledge acquisition and cost of learning transfer. Recent advances in learning tend to replace 'training' with the wider term of 'development' to reflect the more holistic process of learning and the wider choice of learning interventions.

Contents

Disclaimer	ii
Copyright	ii
Abstract	iii
Acknowledgements	v
Glossary	vii
Contents	x
List of tables and charts	xv
Chapter 1: Introduction and Aims	1
1.1 Context and Rationale	4
1.2 Research Aims.....	7
1.3 Profile of the Researcher	9
1.4 Potential Value of the Research.....	11
1.5 Objectives of the Research	12
1.6 Summary	14
Chapter 2: Literature Review	15
2.1 Introduction	15
2.2 Approach	15
2.3 Chapter Structure	18
2.4 Section 1 — What is Training Evaluation?	18
2.4.1 <i>Why do we evaluate and what are we evaluating?.....</i>	<i>20</i>
2.4.2 <i>Conclusion.....</i>	<i>24</i>
2.5 Section 2 — The Kirkpatrick Four Level Framework.....	25
2.5.1 <i>Conclusion.....</i>	<i>30</i>

2.6	Section 3 — Training Evaluation — The Field of Study	30
2.6.1	Group 1 — The Process Focus 'School'	32
2.6.2	Conclusion.....	38
2.7	Group 2 — The Measurement Focus 'School'	38
2.7.1	Conclusion.....	48
2.8	Group 3 — The Outcomes Focus 'School'	48
2.8.1	A challenge in the literature	57
2.8.2	Summary and Conclusions from the Evaluation Literature	59
2.9	Section Four — New Challenges from External Impacts	65
2.9.1	Approach to this section	67
2.9.2	Focus on ideas to build into a new paradigm — Forecasting.....	68
2.9.3	Focus on ideas to build into a new paradigm — Intention.....	74
2.9.4	New Ideas for the Evaluator.....	77
2.9.5	Summary and Conclusion	80
Chapter 3: Methodology, Design and Activity		82
3.1	Introduction & Epistemology.....	82
3.1.1	Epistemology	83
3.2	Research Questions.....	90
3.3	Discussion of Methodology	92
3.4	Research Design Overview	95
3.4.1	Research Phase One — Poll of Attitudes.....	98
3.4.2	Phase One — Research Data Collection.....	102
3.4.3	Phase One — Data analysis methods.....	103
3.4.4	Phase One — Conclusion	105
3.4.5	Research Phase Two — Specialist Practitioner research.....	106
3.4.6	Phase Two — Data Collection	110
3.4.7	Phase Two — Data Analysis	111
3.4.8	Phase Two — Conclusion	112

3.5	Research Phase Three — Formulation of a New Framework — Overview	112
3.5.1	<i>Phase Three 'A' — Formulation of a New Framework — Creation Process</i>	<i>113</i>
3.5.2	<i>Phase Three 'A' — Formulation of a New Framework — Stage 1 — Creation Process — Exposure to new ideas</i>	<i>114</i>
3.5.3	<i>Phase Three 'A' — Formulation of a New Framework — Creation Process — The 'light bulb' moment</i>	<i>116</i>
3.6	Formulation of a New Framework — An Overview of the new Framework — The QED Evaluation Strategy and 4 I's tool-kit	118
3.6.1	<i>Introduction</i>	<i>118</i>
3.6.2	<i>Framework Overview — Approach and Operation</i>	<i>122</i>
3.6.3	<i>Quadrant One — The Subject of the Research</i>	<i>126</i>
3.6.4	<i>Summary of the New Framework</i>	<i>130</i>
3.7	Phase Three 'A' — Formulation of a New Framework — Creation Process — Stage 2 — Testing using a panel of experts	131
3.7.1	<i>Phase Three 'A' Group Benchmark Challenge</i>	<i>134</i>
3.7.2	<i>Phase Three 'A' — Creation of New Framework — Summary</i>	<i>135</i>
3.8	Research Phase Three 'B' — Case Study — Considerations	135
3.8.1	<i>Phase Three 'B' — Analysis</i>	<i>138</i>
3.8.2	<i>Phase Three 'B' Case Study Method — Deployment ...</i>	<i>139</i>
3.8.3	<i>Phase Three 'B' — Approach, Process</i>	<i>142</i>
3.8.4	<i>Phase Three 'B' — Validity Expectations</i>	<i>145</i>
3.8.5	<i>Phase Three 'B' — Measurement Criteria</i>	<i>146</i>
3.8.6	<i>Phase Three 'B' — Case Study Deployment — Data Capture</i>	<i>147</i>
3.8.7	<i>Phase Three 'B' — Main Case Study</i>	<i>148</i>
3.8.8	<i>Data storage and confidentiality protection</i>	<i>150</i>
3.8.9	<i>Summary</i>	<i>151</i>
3.8.10	<i>Research Phase Four — Reaction to Case Studies</i>	<i>151</i>
3.9	Ethical Considerations	152
3.9.1	<i>Summary & Conclusions</i>	<i>161</i>

Chapter 4: Research Project Findings.....	163
4.1 Introduction and Context	163
4.2 Phase One — Results.....	164
4.2.1 Conclusion.....	170
4.3 Phase Two — Specialist Practitioner Research — Results	170
4.3.1 Conclusion.....	184
4.4 Phase Three — Case Study Outputs.....	185
4.4.1 Phase Three — Introduction.....	185
4.4.2 Results from the Retrospective Test	187
4.4.3 Stage Two — Results from Actual Forecasts	189
4.4.4 Initial comparison of transaction costs.....	195
4.5 Phase Four — Overview and Results	197
4.5.1 Stage Four — Reflections.....	200
4.6 Summary of Research Activity	200
Chapter 5: Research Project Discussion / Interpretation	201
5.1 Introduction	201
5.2 The Literature Review — Reflection	201
5.2.1 Primary Questions	203
5.3 Phase One — Attitude Survey	205
5.3.1 Reflection on this Phase	210
5.4 Phase Two — Specialist Practitioner Research	211
5.4.1 L&D Evaluation Practice.....	211
5.4.2 Stakeholder Needs and Outputs	214
5.4.3 L&D Process Improvement.....	216
5.4.4 Reflection on this Phase	219
5.5 Phase Three and Four — Case Study	220
5.5.1 Reflections on this Phase	223
5.6 Conclusions.....	225
5.7 Suggestions for Practice	227

5.8 Suggestions for Future Research.....	229
Chapter 6: Learning Reflection	230
Appendix 1 — Guerçi and Vinante (2011) — Framework	237
Appendix 2 — Richard Griffin (2011) Evaluation Framework....	238
Appendix 3 — Phase One — Survey Questions.....	239
Appendix 3a - Phase Two - Contributor Briefing Document.....	240
Appendix 4 — Phase Two — Survey Questions.....	240
Appendix 5 — Phase Three — Expert Workshop Overview	245
Appendix 6 — Phase Four — Survey Questions	246
Appendix 7 — List of References and Bibliography	247

List of tables and charts

Table or chart	Page
Kirkpatrick model: Levels 1-4	26
Training evaluation: ideas and frameworks categorised	30
Landers and Callan: models of training evaluation	31
Framework for evaluating tools and methods	31
Comparison of Robinson and Carpenter evaluation models	35
Kraiger, Ford, and Salas: categories of learning outcomes	49
The Full-Scope Evaluation Model: four stages	52
Passmore's SOAP-M framework	55
Elements of Passmore's framework and related tools	55
Comparing the Basareb and Thackeray approaches	58
Literature review: questions and answers	65
Trends influencing the learning and development function	66
DiPiazza and Eccles: four methods of prediction	71
Walonick: Forecasting framework	73
Assessment of progress	81
Crotty process	85
Core questions: developing a new evaluation framework	91
Secondary questions: developing a new evaluation framework	92
Mixed methods approach: stages and related actions	93
Saunders et al approach: epistemology and constructivism	94
Overview of Research Phases	95
Research phases: details of design, deployment and analysis	96
Research phase one – poll of attitudes	98
Criteria for testing credibility of sample size (phase one)	101
Steps for content analysis	103
Research phase two – specialist practitioner research	106
Clarification of processes involved in phases one and two	107
Criteria for testing credibility of sample size (phase two)	109
Research phase three – formulation of a new framework	112
Kirton / De Bono: stages in formulating a new framework	113
Existing vs. new evaluation paradigm	117
New Evaluation Framework	118
Traditional Approach to Evaluation	119
Framework driven approach to Evaluation	120
Framework Quadrants and associated Toolkits	121
Creating the evaluation strategy	122
Quadrant one toolkit	127
L&D functions and related factors	128
Quadrant one vs. the traditional / Kirkpatrick approach	129
Testing by panel of experts: weaknesses and mitigation	131
Case-study method: stages and related tasks	136

Yin: Sources of data collection for case studies	137
Strategies to counter reliability problems with case studies	138
Activities for testing the evaluation framework	144
Research phase four – reaction to case studies	151
Steps for maintaining ethical and commercial standards	154
Overview of the four research phases	163
Do you formally evaluate learning in your organisation?	165
Do you use the same approach for all programmes?	165
If yes – to what extent?	166
What technology do you use when evaluating?	166
To what extent does your LMS drive the evaluation you do?	167
If no, why not?	168
How do you know a programme has been effective?	168
How would you describe your attitude to evaluation?	169
How would you describe your attitude to evaluation?	171
Which types of development do you evaluate?	171
Why do you evaluate?	172
Which evaluation methods do you use?	172
What is the data from your evaluation process used for?	173
What decisions are made based on the data in your report?	174
In your view, how are the reports/data regarded by stakeholders?	174
What process do you first operate when a request for training reaches the function?	175
Do you create an outcome target for a programme?	175
Do you identify learning points for each programme?	176
Do you identify post-course actions for the delegates for each programme?	176
I believe that evaluation should be a key factor in the design and delivery of a programme	177
To what extent are you interested in confidence and/or engagement as a course output?	178
What approaches are used to embed learning?	178
What new technologies (e.g. social media etc.) are used to help embed learning?	179
Indicate areas in which managers have had development	180
What percentage of responses do you receive post-course?	180
What methods have you used to stimulate responses (most and least effective)?	181
What methods have you used to foster a culture where the delegates voluntarily evaluate their own learning?	181
How is the L&D function evaluated?	182
How is the budget for development structured?	182
What new approaches to learning are being implemented within the organisation?	183

How are the results of those new approaches being evaluated?	184
How did you learn to evaluate?	184
Steps involved with phases of the research	186
Courses evaluated to test the new framework	187
Retrospective results of testing	187
Company numbers and descriptions for use with charts	189
Organisations' choice of courses for evaluation	190
Organisations' choice of courses for evaluation (with descriptive detail)	191
Expert group forecast expectations	192
L&D group forecast expectations	193
Forecasting expectations vs. actual results	194
Actual forecast average per organisation vs. target	194
Actual forecast average per course vs. target	195
Cost comparison – by course	196
Cost comparison – by organisation	196
Before the experiment, I believed the new method would work	197
What do you think are the benefits of this new approach?	198
What do you think are the risks of this new approach?	198
Do you believe that the cost savings are an acceptable trade-off for the reduced precision of the framework?	198
I now trust this approach	199
We will carry on with this new approach	199
Recap:- Structure of the research	205
Review of core research questions	225
Review of secondary research questions	226
Guerci and Vinante: Framework of stakeholder needs and drivers for training evaluation	237
Richard Griffin Evaluation Framework	238
Expert workshop overview	245

Chapter 1: Introduction and Aims

Since Adam Smith mentioned the usefulness of training in 1776 in the 'Wealth of Nations', training has been regarded as a core organisational activity and one that receives substantial funding from both governments and organisations. In 2013, in the US, the reported annual training expenditure on training alone (not ancillary services or training payroll), was reported to have risen to \$146 billion (ATD 2013), and the level of training investment has risen year on year in the UK since 2008 (CIPD 2015). Therefore, it appears that organisations still believe that training is important and continue to support it through tangible investment.

Whilst the importance of training is recognised, the evaluation of the training impacts or outcomes, which should justify the investment made, is recognised as being much less important. Kontoghiorghe (2004) reported that only 10-15 per cent of training actually transferred into useful and measurable job performance and Griffin (2011), suggested that the paucity and operation of evaluation practice risks the credibility of the L&D function. Even a recent CIPD (2015) report, revealed that 14 per cent of organisations carried out no evaluation at all of their development spending.

The concept of evaluating the outcomes from learning is not new. Arguably FW Taylor (1911) began the formal process as he noted the improvements in productivity in trained staff and their increased effectiveness on production lines. However, possibly the most significant development in L&D evaluation took place across the 1960's as the Four Level Framework conceived by Donald Kirkpatrick (1954) gained prominence and subsequently became ubiquitous in practice. The continued use of the framework has, in effect, inadvertently led to it becoming the core paradigm in training evaluation. It is an interesting (if irrelevant) exercise to consider how little there is in business practice that still continues to be deployed virtually without change from that time (or even from this era).

Since 1954, academics and practitioners have attempted to adapt, shape or refine the Kirkpatrick framework in order to improve it as there is a wide recognition of its limitations. They have included advances in process as well as in the development of core measurement techniques; the most significant being the work of Jack Phillips in applying a Return on Investment calculation to measure financial ROI (Phillips 1989). This has delivered a robust and sensible method to create repeatable (if complicated) financial Return on Investment. However, popularly known as 'Kirkpatrick Level 5', it has helped to reinforce the Kirkpatrick framework, which still retains its primacy in the practice of evaluation in 2016.

This primacy may be a fundamental problem for the L&D function because, in difficult markets, two early victims of corporate spending cuts continue to be Marketing and Training. Therefore, particularly in more uncertain economic periods, it is important that L&D functions have an effective, robust and straightforward method of proving the worth of their (often excellent) programmes to avoid budget reduction. However, L&D continue to suffer frequent cost reduction as insufficient positive evidence exists (other than anecdotal) of their worth, or of being able to prove a return on the investment from L&D using current evaluation practice. Whilst this is still the case in L&D, marketing investment has been transformed through the use of online data and new metrics, processes and approaches to measurement and evaluation. This innovation suggests there is no real reason for L&D not to follow suit, to improve evaluation and protect levels of investment.

There is also plenty of evidence of evaluation taking place in the corporate market place – often a result of the processes contained within Learning Management Systems ('LMS') but, in my experience, there is still a problem in proving business impact, often because the data collected is usually more suited to proving the efficiency gains of the LMS itself rather than the value of the learning. Also, the majority of the LMS providers use the Kirkpatrick framework within their software processes and this simply serves further to embed the framework.

A number of academics and practitioners have attempted to break completely away from the Kirkpatrick concept at various times, (including Cascio (1982), Kraiger (1993), Spitzer (2005), and Brinkerhoff (2003); some with limited success but rarely gaining the level of acceptance needed to challenge the Kirkpatrick primacy. One of the many problems for change is that, when discussing L&D evaluation with line managers (who may have had some formal management development), they will challenge any attempt to stray from the Kirkpatrick model (even when they do not understand what is being offered as an alternative), as it has been part of formal development in (for example), the Chartered Management Institute (CMI) and the Institute of Leadership and Management (ILM) management and leadership programmes. This may also build on the view that change happens when supported by the most respected 'schools of thought' and the dominance of Kirkpatrick (and Phillips) tends to overwhelm the efforts of others to create an alternative approach. Even in the social science research methods literature, the Kirkpatrick (1979) method is the evaluation method of choice; for example, in Gray (2009).

However, it is my contention that the Kirkpatrick paradigm itself stands in the way of advances in evaluation and that a range of new ideas and factors should drive a different method as well as a different discussion of what should actually be evaluated. In fact, many L&D professionals struggle with the idea of 'how to do Kirkpatrick' rather than 'how to do evaluation'. I also submit that the Four Levels have a fundamental problem – and this problem is supported by a majority of evaluation practitioners in the field in that it simply does not deliver the results needed by the L&D function. Despite the legitimacy and traction of the name, the Kirkpatrick method does not actually deliver credible results against the cost invested and value created. Also, other evaluation methods linked to, or built upon, the Kirkpatrick framework are complicated, hugely resource intensive and 'as good as guessing', in the words of one L&D Manager. This bold claim is also reflected in the number of L&D departments operating the Kirkpatrick method, who then

subsequently lose or have their budgets cut, as they are unable to prove or demonstrate their own value.

My original objective was to create a new paradigm to rival that of Kirkpatrick but, as this research shows, the creation of a framework does not automatically mean a paradigm has been created - only the possibility. And, ironically, this research may reinforce the idea that the biggest resistance to change is within the L&D function itself.

Therefore, a more pragmatic approach was chosen to create a rival set of processes and structures than could deliver the L&D function a choice in evaluating. This new approach and its potential acceptance will come only from moving away from the long, slow and expensive approach that is well known, to a cost effective, robust and rapid approach which is less well known but which can create outcomes at the same level of accuracy as the most rigorous summative evaluation processes. The big win for L&D teams is that they can then begin to use evaluation as a strategic decision method, linked to straightforward tool-kits, that will allow them further to enhance credibility over time.

As the dream of creating 'the' new paradigm became more realistic during the research phases and literature review, that realism became the driver to deliver a more focused and targeted method to provide a practical framework for practice.

1.1 Context and Rationale

A useful definition of evaluation provided by Sutton (2006) is to see evaluation as the process utilised by the appropriate function (usually L&D), within the organisation to identify and measure the results or outcomes (and value) created from investment in the training courses or programmes delivered, purchased, or managed by the L&D function.

An early decision was made to focus this research on 'training' and development rather than the whole field of L&D. This means that

workshops, blended learning, e-Learning and other forms of training are included, but that the subjects of coaching, facilitation, mediation, events and consultancy are not included in this research, despite the huge overlaps in the management and evaluation of these processes by the L&D function. In my practice, these other areas of development, including coaching are evaluated, and some commentary does appear in this document but only as passing remarks.

One of the challenges with a subject such as evaluation is that it straddles a broad range of theories, concepts, approaches and practice. In addition to the academic perspective, learning evaluation sits within a complex number of L&D processes that are part of Human Resources ('HR') and organisational processes. One of the potential advantages that a scholar-practitioner can bring to the subject is to properly contextualise the subject and to ensure that some clarity can be found within the complexity of competing positions. This may help to situate any new ideas within a logical context in order to help the practitioner determine appropriate choices for future evaluation.

Depending on the size of an organisation, the L&D function is often part of an HR function. The Ulrich framework (1988), built on the theory of Human Resource Management, (HRM) changed the fundamental focus of the Personnel and Training departments to that of an HR function. This new function focussed on developing the management community to operate more effectively, allowing the HR function to split its time between a consultative approach supported by effective HR service delivery. In effect, this means that the L&D function within HR has often become part of the service delivery chain and is often disconnected from the processes of Organisational Design and Development ('OD') of which it was a part before Ulrich. Whilst evaluation could be part of a strategic proposition for both HR and L&D, a range of CIPD reports (2010, 2014, 2015) suggest it has not achieved that status. This leaves both functions potentially vulnerable to operational, financial and investment priorities, especially when investment becomes restricted.

Within the wider HR remit sit bodies of knowledge that affect L&D. They include Human Capital; Talent Management; Reward; Culture; Engagement; Change Management; Compliance etc. and each of these has its own bodies of knowledge as well as focus and interest, depending on the culture and structure of the organisation and its HR function. It is also the case that the HR function uses technology to handle the increasing data load from the organisation as well as the many technology demands required those bodies of knowledge. This reliance on the need for HR information often affects the technology spend available for L&D and the level of investment in any LMS.

A focus on the training measure of 'cost per head' can also drive L&D into providing greater e-Learning on the basis of efficiency. External technology suppliers have changed the landscape in technology driven learning and the level of innovation and change is exhibited annually at the Learning and Technology Show in London. The rise of innovation in Hardware, Software, Apps and Virtual Learning is reflected in the show notes and guides. Those guides also illustrate the paucity of evaluation innovation in the technology space. In fact, the two main providers still use the Kirkpatrick framework as their underpinning concept.

This lack of innovation and effectiveness has also led to the failure of a number of evaluation practitioners and specialist evaluation technology suppliers over the last three years. As one evaluator reported "L&D are just not interested enough to invest in evaluation – and no one seems to be asking them to prove value, despite the increasing sums spent on training and development".

L&D itself is also becoming increasingly busy in order to meet its objectives in a resource efficient manner. In a function that is well resourced, according to the CIPD, one would expect to see the following processes as minimum offering. (CIPD 2014).

- Learning Needs Analysis
- Learning Design

- Learning Delivery (utilising appropriate channels)
- Learning Review
- Learning Admin
- Learning Evaluation
- Supplier Management

Therefore the CIPD now fully acknowledges evaluation as a core requirement of L&D and has it as a subject within its professional development. However, the core text for the syllabus is the Kirkpatrick framework although there is now limited recognition of some other approaches, particularly the Kearns (2002), the Success Case Method (2002) and Phillips (1989).

Within these L&D processes and their associated bodies of knowledge sits a range of technology and learning solutions, concepts and platforms for delivery, with recent innovations including 70:20:10; Gamification; Virtual Reality; Neuroscience and associated advances applied to learning (e.g. NeuroLeadership); Big Data and analytics; Social Media, as well as new platforms of engagement. Within this level of innovation, traditional problems in improving practice in ideas relating to memory, learning, behaviour, motivation etc. are also being affected by new ideas and concepts. In effect, the L&D practitioner has to be adept at keeping up with a range of competing demands and advances in technology, learning and practice in addition to the demands of working in an organisation. It is understandable therefore to see that evaluation may become less important if it is too difficult or too resource intensive, especially if there is little demand to prove value or return.

1.2 Research Aims

Whilst I began the research with a profound belief that well-embedded training adds tangible value through the implementation of 'practical knowledge' Anscombe (1959), I recognise that I had to put aside those values for the duration of this project to allow me to examine the problem 'afresh' and to be objective in both the definition and conduct of the research.

In order to focus my research, many of the core training processes would be out of scope for this project to maintain focus on the areas of evaluation. Some had to be reconsidered later in the process as, in developing the new framework, this highlighted the need to improve these core processes to be able to evaluate quickly and robustly.

Also, at the beginning of the research, examples of areas of interest that which were out of scope of the project included:

- The actual range of delivery mechanisms chosen by the training provider
- Any evaluation of the purchasing decisions regarding the training provider, venues or facilities
- The use of any LMS
- The nature of Training Needs Analysis or Assessment
- Any evaluation of Human Capital or Talent Management processes linking business strategy to human performance
- Any commentary regarding the choice of company performance management systems
- The structure or human capital processes within the training function
- Any commentary on the competence of trainers

The aims of the research were therefore threefold:

- To create a new approach to the evaluation process (how evaluation is conducted)
- To construct a new approach to evaluation measurement (how learning impact is measured)
- To demonstrate the accuracy of both against current methods

To summarise, the rationale for the research developed to create an evaluation framework to enable the function to prove its value through a new combination of strategy, process and metrics.

1.3 Profile of the Researcher

I currently work within a specialist evaluation practice that helps L&D prove their value from their programmes and courses measured against the overall budget. My career journey to this point has been a series of steps that has allowed me to gain the relevant experience to understand the challenges and opportunities for this subject in the commercial world.

My initial training was that of a professional musician giving the understanding and experience of 'performance', including its psychology and impact, and this was translated into skills as a trainer and consultant. The next stage of my career was involved in Sales and Marketing and I developed personal skills in both the practice and psychology of both. Marketing triggered the realisation of the need for robust data as a means to prove 'softer' concepts to more pragmatic and numerate sales teams within which the marketing department often operated. The final phase of my career moved from operational management into consultancy, through a number of roles in which I gained experience, including:

Role	Experience
Director of Sales & Marketing within Automotive Manufacturing	Gaining strategic perspective and the need to justify budgets
Head of Training for an insurance firm	Gaining process and management expertise
Interim CEO for a law firm	Gaining strategic focus
Operations Director for a consultancy	Gaining the skills of facilitation, training and consultancy and building skills and qualifications in change management, psychology and business administration
Entrepreneur/consultant	Gaining the skills of running, growing and selling a large Human Capital business that included activities in training, assessment and evaluation

As a result, I have a track record in people development, evaluation and measurement, ranging from individual assessments for Private Equity

firms investing in mergers and acquisitions, to the evaluation of corporate functions. Also, I have a track record in entrepreneurial activity, operational management, training delivery and product delivery that lends itself particularly well to the subject of evaluation. I have been part of a team that launched Emotional Intelligence into the UK market, and created and sold an e-Learning organisation. I have become 'battle hardened' in the subject of people development and evaluation by extensive execution in a wide range of internal business disciplines, as well as client delivery through working on many high profile change programmes.

In 2002, I created a company that grew to become a premier provider of executive leadership and management programmes in tandem with executive personal development and coaching. However, it does not take long within a training company to realise that the work delivered is constantly at risk of being cancelled if the L&D function cannot evaluate its effectiveness. From that grew my current interest and change in focus to a specialist evaluation provider, created from a division within that training company. I therefore transitioned into focussing on evaluation as a sole discipline and recently launched a specialist company to work with clients and deliver value from evaluation. Part of my practice is to help develop, consult with, or coach people within a L&D function to be more effective at proving their value as well as viewing evaluation from a different, positive perspective and this research has proceeded in tandem with many of the areas of innovation within my own organisation. We are described by those that meet us as being unique in our focus and versatility as specialist evaluators in L&D. I have also achieved a number of business, functional and academic qualifications that have resulted in a wide range of 'soft skills', tools, and frameworks.

I therefore considered myself to be ready and credible enough to tackle a DProf process to attempt to achieve the objectives I had set for myself, based upon the delivery of successful projects and transformation of evaluation in a large number of client organisations.

1.4 Potential Value of the Research

There is a twin track of activity involved in this research process. My aim is to replace the Kirkpatrick framework with something more useful and relevant for L&D and, in order to achieve this, I will need to create something new. Working to generate ideas and to indulge my passion for learning is the second area of activity in the research and the merging of these tracks will create a satisfactory personal conclusion. In considering the idea of wider value creation from this research in evaluation, individual belief and enthusiasm have to be tempered with the realities of the day-to-day practitioner interest in the subject. However, I see strong benefits to a range of parties as a result of this research.

To those in the field, my belief is that the survey of the practitioners in the report will help create optimism and opportunity in equal parts, but also that the focus on the practical aspect of implementation and approach, rather than simply a theoretical focus will help challenge and identify opportunities for new methods as much as to move forward knowledge in the field of study.

The research will generate a number of products, books, white papers, software and training interventions for practitioners. I believe that the dissemination of knowledge to those that want to know, and are interested in trying out innovative approaches is best served by this approach. The L&D function is hugely influenced by external, valid research and the nature of this qualification, in blending professional practice with theoretical rigour, should add value in building both the confidence and legitimacy of the outcomes.

Additionally, a robust and thorough review of the evaluation literature, perhaps with a different perspective, can find fresh meanings, themes and perspectives that can be blended with wider reading to create opportunities for wider research.

1.5 Objectives of the Research

At the beginning of the research process, the initial thinking around this document was to create a new conceptual framework of training evaluation, particularly with regard to larger scale management development programmes. As the process continued, the idea of restricting the research to management development programmes was removed, as, for a framework to be effective and useful, it must be relevant over the entire L&D function also, a more pragmatic approach drove the recognition that the ability to create a paradigm is outside of the control of the researcher.

Therefore, the objectives were amended to create a new evaluation framework that would give people involved in the L&D function a choice when they evaluate, and which delivers straightforward, robust, credible and cost effective outputs. As previously stated, the objectives for the research were:

- To create a new approach to the evaluation process (how evaluation is conducted)
- To construct a new approach to evaluation measurement (how learning impact is measured)
- To demonstrate the accuracy of both against current methods

In order to assist with the initial thinking, a series of questions was posed at that time to help shape my initial thinking to create the next steps in delivering the research objectives. They were:

- What attitudes exist towards the subject of evaluation in the L&D practitioner base and how might those attitudes affect any change of approach?
- How are new ideas and approaches in the design and delivery of training and development being reflected in advances in evaluation practice?
- Should evaluators consider a wider range of intangible aspects, such as confidence and morale, in developing evaluation metrics?

- What role could the concept of 'Decision Theory' have in adding value to evaluation perspectives and measures?
- Could the 'embedding' of learning affect the outcomes from courses and, if so, how can successful learning 'embedding' be accomplished?
- How can a wider range of the stakeholders in learning (for example, delegates) become more actively engaged in evaluation to reduce the load on the L&D function? To what extent would this be a simple change of process, or a more complicated change of culture?
- What areas of practice and literature could present opportunities to generate a new perspective as a practitioner and researcher in developing a new framework?
- What new approaches might exist in the literature that could be reinvestigated or re-imagined to create evaluation metrics to support a new framework?
- What techniques from the creative thinking and innovation tool-kits could shift the entire 'world view' of evaluation?

This list of questions was refined and reshaped by more relevant and penetrating questions – partly to respond to the challenges posed by the literature, but also by the realities of work and life ending and became:

- Why is the Kirkpatrick framework problematic to implement?
- Why are the results from evaluation not always trusted or valued?
- Why is the evaluation process not seen as a vital part of the L&D function, given the need to prove results and value?
- Why is there so much academic discontent about the framework, yet it still remains as the normal means of conducting evaluation?
- Why is it still relevant and legitimate in a world so different from the time of its creation?
- Does this longevity represent a sign of usefulness or not?

The project itself had a number of phases:

The project consisted of a review of the literature that helped develop perspective and generate ideas. A survey process was deployed to examine broad attitudes to evaluation as well as the views of a range of practitioners to investigate current practice – this prompted ideas, themes and learning, that drove my ability to create the framework as well as add value and context to my own learning. A new framework was created using outputs from my personal reflection as well as from synthesising ideas from research and the literature. The framework became part of a case study approach and the accuracy of the results was tested against the expectation of the case study participants.

1.6 Summary

My motivations for conducting this research are therefore:

- Functional: to help build the credibility of the L&D function to prove value
- Personal: to stretch my knowledge in two ways, the first as an academic operating at an appropriate standard and achieve a level of academic recognition for research and, secondly, to build practical knowledge to apply in a commercial setting
- Practical: to build something that can be validated to achieve the first two motivations

I am confident that my journey so far in my personal and organisational life can become the foundation for the creation of something meaningful. Whether that creation can achieve sufficient traction in the real world to allow it to become useful will remain to be seen.

Chapter 2: Literature Review

2.1 Introduction

This chapter will identify the approach to the literature review as well as decisions made that shaped the nature and process of the review itself. The review will discuss the 'schools of thought' within the evaluation literature, as well as specific concepts from outside the evaluation literature that have guided and shaped my own research as well as that of practitioners and other researchers in the field of evaluation.

I initially identified a number of questions as a practitioner that have shaped this research to achieve the objectives. They were:

- Why is the Kirkpatrick framework problematic to implement?
- Why are the results from evaluation not always trusted or valued?
- Why is the evaluation process not seen as a vital part of the L&D function, given the need to prove results and value?
- Why is there so much academic discontent about the framework, yet it still remains as the normal means of conducting evaluation?
- Why is it still relevant and legitimate in a world so different from the time of its creation?
- Does this longevity represent a sign of usefulness or not?

2.2 Approach

A literature review on the subject of evaluation in this field has challenges because of the large literature associated with the subject of learning, as well as the processes of learning, training and wider development. Evaluation has also been applied in other fields, such as marketing, projects, social impact etc. and this risks making a review of the literature overly large or diverse. Therefore, a specific strategy needed to be adopted using a phased approach to the literature using the analogy described by Hart (1998) in 'camping' in closely aligned texts to the main subject and 'trawling' through a wide range of associated literature in

order to synthesise wider learning and create new interpretations of meaning.

Phase One of the literature review consisted of 'camping' in the existing learning evaluation literature, beginning with the Kirkpatrick framework and forming a critique of its effectiveness. This then informed the subsequent examination of the progression of thought and practice within the subject, including identifying the key 'alternative' thinkers in the evaluation literature including Boudreau (1983), Spitzer (1984) and Kraiger (1993). As an extension to that process, more recent areas of evaluation practice that used different methods or approaches would also need to be examined, particularly the work of Basareb (2007) and Griffin (2011).

Phase Two of the literature review involved 'trawling' through a range of associated fields of literature. Initially, this focused on functions that also had a need for evaluation, or where the practice of evaluation was more well-established. In examining such a wide body of literature, the criteria of 'relevancy' and 'transferability of approach' to practice were used to drive choices. The practise of evaluation were examined in Marketing (often another area of vulnerability to cost-cutting in organisations), which led to the discovery of some key metrics including 'Net Promoter Score' and 'Perceived Value'. However, the fields of literature in Social Impact Research, Economics and Investing were briefly investigated but discarded as they failed to meet the defined criteria.

The function of L&D is also involved in the processes of assessment, design and delivery of training and an early decision was taken to discard these avenues of research even though there is interdependency of process with evaluation activity. This decision was taken to give adequate time to focus on the different bodies of literature outside of L&D, to avoid being driven too much around existing and conventional L&D practice. However, one area of L&D process did emerge from the wider review as being significant and which had synergy with evaluation and this was Learning Transfer, with some consideration given to the work of Holton et

al (2000). Interestingly, much of the real innovation in thinking came from wider social reading, but using the structure of this review helped integrate both formal and informal avenues of learning.

Phase Three of the literature review consisted of 'camping' in areas that were determined would be the basis of a specific approach. These were deliberate choices that had emerged from the literature and appeared to be the basis of a new framework and meeting the needs of the objectives and included: forecasting; 'intention'; metric development; the Law of Large Numbers and Decision Theory. These areas then formed the basis of specific tool-kits that drove the approach to the Case Study discussed in Phase Three of the research.

In determining an initial strategy to help structure the research, the guidance of Santos et al (2006) proved to be of some initial, if limited use. Their contention that a researcher should pay heed to traditional bibliographic principles including, for example, citation analysis initially proved to be attractive, however, deploying the more pragmatic approach of selecting the key thinkers in the field and tracing their own influencers and followers, produced a more interesting blend of traditional as well as fresh thinking in the field. It was that fresh thinking that added some of the most significant insight and meaning to the research. As a final point of reference I needed to remain aware of the structure of a good literature review as shown in Gray (2009), which identifies a range of categories and criteria for assessment of the literature review itself. These include the assessment of:

- Coverage
- Synthesis
- Methodology
- Significance
- Rhetoric

A brief review of this concept is carried out at the end of this chapter to assess whether the fundamentals of the review met the requirements that Gray recommended.

2.3 Chapter Structure

<i>Section 1</i>	An outline of the concept of evaluation and how the nature of the subject itself creates complexity for practitioners
<i>Section 2</i>	A critique of the core Kirkpatrick framework and its position in the field of evaluation. Questions are raised regarding its suitability for the modern world
<i>Section 3</i>	A review of a selection the key thinkers in the field of training evaluation. This section is subdivided into three further groupings to illustrate 'schools of thought'. This section also outlines some of the more technical areas for debate
<i>Section 4</i>	A discussion of a number of the challenges facing the modern world for the L&D practitioner and the extent to which the prevailing paradigm adds value (or not)
<i>Section 5</i>	An investigation of areas of interest for my own research to be tested and reported upon in a later chapter
<i>Section 6</i>	An outline of some areas for other practitioners to consider as the field continues to develop – some of these areas may also be considered in my own research

2.4 Section 1 – What is Training Evaluation?

As previously stated, a useful definition of evaluation is provided by Sutton (2006) to see evaluation as the process utilised by the appropriate function (usually L&D) to identify and measure the results or outcomes (and value) created from investment in the training courses or programmes delivered, purchased, or managed by the L&D function.

However, in reality, the picture is somewhat less clear. From Adam Smith to recent times, a range of organisations, governments and thinkers acknowledge the value of training by using a range of economic indicators. According to Griffin (2011), the UK government states that skills and training contribute "to employment, productivity and growth" (Department for Business, Innovation and Skills, 2010, p. 5) and even in the face of the recession from 2007 – 2012 in the US and UK, organisations in England spent some £37 billion annually on workforce training (LSC, 2009) and in the US the figure was over \$130 billion (ASTD, 2009).

However, this confidence and investment in training becomes less clear when applied to individual organisations and especially when attempting to determine the value of training to the organisation itself or to specific individuals. Plant et al (1992) have identified the key processes in the body of work in Human Capital Accounting which has attempted to link the intangible value of people to the organisational 'bottom line' through capitalising 'human assets' and the value of training as have (Likert 1961), Flamholtz (1969), Hermanson (1964), Cascio (1982). However, Griffin (2011) reports that the paucity of evaluation expertise hinders efforts in Human Capital Accounting as evaluation has insufficient 'legitimacy or consistency' of application to add sufficiently robust information.

Evan a cursory examination of the HR or L&D trade press finds assertions that training delivers value; however, some of this confidence was challenged by Guyott (1974) who identified a 'Hawthorn effect' in control groups that showed no performance gains between groups who had been split into 'intention to train' and contrasted with those that had actually been trained. Whilst a single study does not constitute evidence in itself, there is increasing evidence of practitioner research emerging in L&D membership communities such as TrainingZone.com (2016), where a recent headline pronounced that 'Training does not work – nor does it create any inherent value in itself'.

The widely held belief that that training is effective could be negatively affected by the processes and concepts used in evaluation. It is also the contention of those within the Learning Transfer 'school', including Holton et al. (2005), that ineffective learning transfer actively negates the positive effects of training delivery and that evaluation has to differentiate between the processes of training delivery and learning transfer in order to determine impact or to be able to add value to the whole learning process. These themes will be developed further in this chapter; however, a thought for consideration must be whether, if neither 'evaluation' nor 'transfer' are operating effectively within an organisation, there is much point in training people at all? This view

reinforces Griffin's (2011) suggestion that the L&D function has too little credibility because evaluation needs to be improved in structure, concept, and process. Further discussion will address these areas later in the chapter.

There are other factors underpinning evaluation that also may begin to explain some of the issues about the effectiveness of learning evaluation. Hashim (2001) argues that the actual motives for evaluation are often 'ignored or approached in an unconvincing or disorganised manner' and Gutek (1988) criticises the poor systematic process except for the application of the Kirkpatrick method. Brinkerhoff (1988) argues that this lack of a logical approach negatively affects the ability to collect and interpret meaningful data at all. In addition, Tasca et al.(2010) have suggested that budgetary constraints on the L&D department have led to the adoption of simplistic IT systems for evaluation because of a failure to generate a cost/benefit from the actual evaluation process itself, and Scourtoudis and Dyke (2007) reinforce this view claiming that barely one per cent of organisations actually measure the impact of learning on performance when using a LMS. This level of ambiguity and criticism of the process and practice of evaluation creates an issue. Whilst there is a tacit acceptance for the need for training at a macro level, there exists, at the micro level, a range of issues and factors that face practitioners and need to be addressed to ensure that the wider view of training and development is based on solid evidence and sound practice.

2.4.1 *Why do we evaluate and what are we evaluating?*

The Kirkpatrick (1954) framework creates an evaluation approach that focuses on four main levels, including: delegate reaction; knowledge acquisition; behaviour change, and organisation 'results'. The framework is both simple and simplistic, as it considers the needs of the organisation in its construction, with the learner as a (passive) recipient of the learning. However, there are many reasons to develop people and these can become increasingly complex, depending on the needs of the many parties involved in the learning process. The evaluation process should be

capable of matching those needs and the complexity of the learning context.

Guerci and Vinante (2011) created a useful framework of the various stakeholder needs and drivers as well as the vested interests that this complexity generates. The framework considers the needs of the organisation, the internal training team, the external training partners, and the delegates. The framework, by allocating benefits to process outcomes, reinforces the risk that in most scenarios it is the interests of the external training organisation that are often better served than the organisation. The irony of this conclusion cannot be stressed enough, particularly when the external training organisation then is required to carry out the evaluation of their own work. This framework can be found in Appendix 1.

The learner (often referred to as the delegate) is a key stakeholder in the process of development and, it can be argued, that they could and should be more involved in the evaluation process. Ibrahim (2008) outlines a range of motivations for learning ranging from an individual desire for improvement, through to desired changes in work performance, as well as development in concepts, skills and attitude. In each case surely this should create an opportunity for learner-driven evaluation. In this approach impact is measured by changes in job performance although, as pointed out by Bramley and Kitson (1994), this is always more appropriate when the learner has some control over the learning process and delivery of outcomes, and is not just expected to 'supply the motivation', but this is not always the case.

Geertshuis et al. (2002) are noted for their support for learner-based evaluation and also contend that the whole evaluation process sits within an evolving construct where learners are becoming 'customers' and that the societal factors are becoming increasingly 'multiplicative'. This leads to an issue for the evaluator in contending with a process handling 'multiple variables and multiple interdependencies.' However, to contrast with this view, Billett (2007) contends that, at the simplest level,

workplaces provide opportunity to learn and learners supply the motivation – and these are the only components that can and should be evaluated. However, this approach would depend on measures of delegates', (or customer) 'satisfaction' and Lingham et al. (2006) challenge the concept, contending that the effectiveness of training should be based on the ability of the learner to apply the knowledge, skills and attitudes obtained in the training course as 'relying on satisfaction measures will never deliver this outcome'.

Whilst the motivations and desires of the learner are drivers for effective training, Griffin (2011) contends that more effective 'learning triggers' than individual 'employee self reflection', are the organisational triggers of 'responding to policy developments or incidents' and 'statutory requirements', as well as the formal HR or L&D process of Training Needs Analysis that often drives the core L&D 'offer' of development. His contention is that organisation investment drives the need for the creation of value primarily for the organisation, through both direct and indirect application of the learning in the workplace.

Kirkpatrick (1954.1979), and followers such as Phillips (1999), also focus on the needs of the organisation when evaluating viewing the L&D department as being at the heart of the development process rather than the learner. This organisation-centric view is by far the strongest across the body of literature and, paradoxically, whilst it is the learner that 'does the learning', it is the L&D function that is responsible (and usually accountable), for the effects of the learning, with the learner often absolved from any responsibility in the evaluation process or in the transfer of learning. Cascio (1992) also builds on the specific need within the organisation in considering the requirements of the sponsoring 'line manager' as part of the process and suggests their requirements also need to be part of any training design. Whilst Giangreco et al. (2010) also agree with the organisational focus, they propose a different approach by focusing on Utility perception and Learning Transfer, requiring all of the different parties in the learning to play their respective parts.

As well as the complexity of the competing and conflicting workplace needs and stakeholder drivers that drive requirements in the L&D 'offer', there are also wider conceptual and philosophical complexities that can be found in the literature. The extent to which they affect and influence theories of evaluation may create even further confusion for the practitioner.

The economic approach to evaluation has been influenced by Becker's (1975) Human Capital Theory. This approach investigated general and specific training and the extent to which it applies to the sponsoring organisation. Badescu and Loi (2010) have recently pointed out that this creates a specific challenge as *'since it is hard to measure the content of training, inferences are made using information on tenure, quits, turnover rates and earnings growth'*. While this economic approach is prevalent, Heyes (2000) represents a more sociological approach, moving away from large-scale economic indicators and preferring to focus on 'power, conflict and control' that shape learning in the workplace. This, 'softer' approach is developed by Billett and Sommerville (2004), proposing that tangible measures should be abandoned in favour of the creation of 'storylines' and social discourse regarding the take-up of appropriate learning.

Another theme within the evaluation literature suggests that evaluation could, and should, be part of Decision Theory. In other words, providing information to help 'make a decision' as the driver for that information, for example justifying spending to the line, to help decide on the levels of future investment as suggested by Fitz-Enz (1988) and Flamholtz (1985), or in creating information to generate learning inputs and outcomes, based on the continuous improvement of the training offer and individual course effectiveness, as suggested by Flamholtz, (1984) and Cascio, (1991). According to Steele et al. (2015), Decision Theory is usually broken into two drivers - one that concerns the ability to make the 'best' decisions (normative), and the other that explains how people actually 'make' decisions (descriptive). It is the normative branch that is referred to within in this document. The basic principle according to Hubbard

(2007) is to begin with a decision and then choose an appropriate process to determine how to make that decision. Whether that process simply serves to create a probability for the decision to be correct or to reduce uncertainty about the outcome, the starting place is always to define the problem in hand. Hubbard (2014) also states that, as part of any decision, a process should be used to calculate that the value of the decision must be worth the cost of the decision process. This approach is in contrast with the Kirkpatrick framework that can only address the problems it was created to address, not the wider range of decisions that a practitioner may seek to solve. However, as Boudreau (1991) cautions, Decision Theory can be corrupted by decision bias and persuasion theory, with evaluation outputs incorrectly positioned to suit the needs of the reporting party.

However, Griffin's (2011) contention is that low 'organisational operation' of evaluation is the primary negative factor in this confusion and that, as this is really only an emergent field, that 'the lack of a body of knowledge or professional journal illustrates the inherent confusion in approaches'. Aguinis and Kraiger, (2009) also point out the lack of meta-research into new evaluation measures and that differences in approaches are compounded by the 'disparate range of academic disciplines'. For the practitioner, the variety of perspectives, approaches, stakeholders and academic perspectives, allied with a lower level of attention and competence within evaluation practice itself, appears to generate a level of complexity and difficulty that leads to the confusion about what evaluation is, what it is for, and how it should work. An earlier assertion that problems in evaluation may hide whether the process of training works or not appears to be a distinct possibility.

2.4.2 Conclusion

This section aims to illustrate that L&D evaluation exists within a range of complex priorities and interests. The challenge for the L&D practitioner is to be able to deliver the various learning needs of stakeholders, as well as to be able to become more proficient in evaluating the returns within

each of the stakeholder groups. As the demand and investment in training and development appears to remain important at a macro level, the people within the L&D function are becoming increasingly busy and may become reluctant to carry out activities that they may consider to be less important. Also, the apparent lack of a simple solution, other than the Kirkpatrick framework, may limit the desire and opportunity to invest in an area of training practice that also requires specialist skills and processes.

The literature review will now consider the prevailing 'paradigm' and other evaluation solutions, along with some consideration of the ability of a practitioner to be able to operate a new approach within this complex context.

2.5 Section 2 – The Kirkpatrick Four Level Framework

According to Sutton (2006), in 1954 Donald Kirkpatrick created, as part of a PhD dissertation, his idea of a four level causality chain, which linked training to organisational impact. The framework arrived in the wider learning literature later in that decade, particularly with a series of four published articles in the American training trade press. The concept spread by word of mouth and through a wide range of L&D literature, to effectively become a paradigm. It is difficult in 2016 to find learning texts that do not cite Kirkpatrick as the main approach to evaluation and this often includes social research methods for academics, for example, Gray (2009).

Similar to the idea of a 'Hoover' being an archetypal carpet cleaner, the 'Kirkpatrick' model has become what evaluation means to many L&D practitioners. Commentary on practitioner forums such as LinkedIn or within the CIPD discuss how to 'do Kirkpatrick', as opposed to how to 'do evaluation'. The irony is that few really understand what the idea was, is, and was meant to be and what it has become. Let it suffice to say that this one idea has both driven and, arguably, blocked further approaches to evaluation to the present day.

Whilst the limitations of the Kirkpatrick levels are many and frequently recognised Aguinis and Kraiger (2009) argue that, even though the approach is not scientifically based, the level of vested interests will make any shift away from the Kirkpatrick idea difficult and onerous.

The fundamental idea of the Kirkpatrick framework is that it contains 4 levels that, when proven, can be said to have created a robust evaluation, although Donald Kirkpatrick later said in 1994, that this idea was not part of the original aim...nor was it intended to become a form of workplace evaluation – even though it is used as such.

The levels are as follows:

<i>Level 1:</i> Reaction	How the learners reacted to the training messages, environment etc. (Measured by post-course questions – often referred to as 'happy sheets' or, more recently, as 'Smile Sheets')
<i>Level 2:</i> Learning	The increase in knowledge capability (measured by testing)
<i>Level 3:</i> Behaviour	Level of change (measured back in the workplace – though 'how to do this' is never specified). Later papers (1997) suggest the use of control groups
<i>Level 4:</i> Results	This stage links together the effect of the training on results (again, 'how' to do this is not specified – though control groups are suggested in later papers (1978)

In conversations with practitioners they will often quote “we do level 1 & 2 but levels 3 & 4 are such a challenge”, as if they are attempting to deliver a process. Usually conceptually deployed in this way, the first two levels are the most straightforward and within the grasp of the L&D function, so completed most regularly. Levels 3 & 4 are more akin to formal summative evaluation and therefore more difficult so often ignored. Cheng and Hampson (2008) argue that this lack of regular execution of the whole framework leads to a lack of understanding about the most difficult levels. Bramley and Kitson (1994) also argue that the overall lack of data and case studies means there is not enough cumulative evidence in benefitting from these last two levels. In the 1999 – 2003 ASTD Benchmarking Service, the figures show that, whilst around

80 per cent of organisations conducted Level 1 assessment, this percentage fell rapidly as the Kirkpatrick levels progressed and, on average, only about 8 per cent of all training organisations who subscribed to the ASTD benchmarking service made any attempt to ascertain performance or results-based evidence at Level 4.

Bates (2004) summarised a range of concerns with the framework and is supported by Guerri et al. (2010), who stress that the lack of contextual information means any data collected will be over simplified. Also that the causal links between each of the levels imply that progress to the top levels is affected by a lack of progress lower down. However, Alliger and Janack (1989) doubt whether this causality was ever the intention behind the framework and criticise the operators of the framework in abbreviating it in that way.

Two more recent thinkers also tend to dismiss the paradigm in its entirety. Holton (1996) is particularly critical, judging the framework as a 'taxonomy of outcomes' rather than a framework – more 'a mode of thinking about evaluation rather than a method or approach'. He states that the framework also ignores a wide range of problems, in particular that it implies that performance during training is a prediction of post-training performance or that satisfaction translates into learning, and disputes that this is actually the case. Holton (1989) also suggests that each of the levels are strategies in themselves, rather than an overall process, and Kearns (2005) stresses that any approach to evaluation that does not have a means of establishing relevant baseline of competence (etc.) cannot be fit for purpose.

Plant et al. (1992) present the evidence from a questionnaire completed by 600 organisations (each with a workforce greater than 500 people), in which none was able to prove any Level 4 financial benefit, neither did they have a systematic approach to Levels 2 & 3. Plant et al. (1992) stresses that the need at Level 4 is to concentrate evaluation on work skills through the development of work skill taxonomies. In addition, they suggest replacing Level 1 with a more useful reflection process using, for

example, Kelly's repertory grid process. Fitz-enz (1988) also agrees with this approach and created tool-kits to work around the limitations of the approach, whilst maintaining the integrity of the Kirkpatrick framework. This reflects a trend of building on the framework in order to 'fix it'. Sutton (2006) highlights the work of the ASTD benchmarking reports to support his view that "the most widely accepted and used framework for training evaluation consistently fails to produce the very evidence that it purports is essential to demonstrate, namely, value"

As well as challenges to the whole framework, there are also challenges identified with each stage. Darby (2006) suggests that Level 1 is negatively affected by peer pressure, or the position of the evaluation process in the training day (usually at the end – just as people are attempting to leave). Hamblin (1974) and Bramley and Kitson (1994) also maintain there is too great a focus on happiness or satisfaction, leading to the creation of the 'Smile Sheets' approach. Dixon (1990) also stresses that a 'reaction' approach does little more than measure 'enjoyment' and often negates the focus of the training. Hall (2003) goes further and stresses that Level 1 is no more than a means to capture feelings and agree with Noe (1986) that there is no significant correlation between learning and delegate satisfaction and that the framework should not imply that there is.

The assessment of learning in Level 2 can be problematic. Kraiger et al. (1993) challenge the use of the same assessment tools being used to measure skills and facts without differentiation. Airasian & Miranda (2002) suggest that a revised Bloom's Taxonomy should be used to enable this. Lee (2007), also suggested a fix to the quality of learning objectives stating that a lack in this area fails to build on the recent work in defining 'learning factors' utilising: Constructivism, Underhill (2006); Metacognition, Foster et al. (2003), and self regulated learning Boekaerts (1993). These learning factors must be used in any debate about the establishment of knowledge gain, and would give greater legitimacy to the Level. Because of the work in the development of competency frameworks, this could be a useful point – however few competency

frameworks have received much rigorous academic scrutiny to see whether they actually deliver value. Kirkpatrick (1977) himself stressed the need to move away from self-referenced reports and move to a move 'proof based' approach using control groups etc., to check that the training had actually generated the shift in learning.

In Level 3, Kirkpatrick (1979) also suggested that the best way to measure behaviour was to be as pragmatic as possible...."Let's shoot for proof but be satisfied with evidence. In most cases, our superiors would be more than satisfied with evidence, particularly in terms of behavior or results". In many ways this level has proved to be the most problematical in terms of a practical measure, until the work of Cascio and Boudreau (1993, 2002) linked behaviour with Utility Theory.

James and Roffe (2000) state issues with the legitimacy of Levels 3 and 4, as they maintain there are too many variables (from personal motivation to culture), to be able to use this approach to attribute impact to the training. Tyson and Ward (2004) agree with this thinking and interestingly, since the death of Donald Kirkpatrick, his company has begun to rebrand the framework – moving to the descriptive term 'Return on Expectation', and adding in a base-lining stage; more akin to the work of Paul Kearns. However, James and Roffe (2000) have reported that summative survey-based satisfaction measures, which are, easily the most common method this part of the process of evaluation, do not correlate strongly with learning or the transfer of learned skills to the job (Brown (2005)).

One of the challenges for the evaluator is the use of Learning Management Systems by the L&D function. Within each of these systems is embedded the process they have determined is best for evaluation. Of the over twenty LMS platforms identified in the 2016 white paper, 17 use the Kirkpatrick framework as their starting point. CIPD (2014) & Training Solutions (2016). The problem with this approach is that evaluation becomes a method to prove a reduced cost within the wider L&D budget; a point used to market and sell the LMS offering.

2.5.1 Conclusion

It is an interesting reflection that, once a paradigm, meme or example of the Lindy effect appears, even the protestations of the author or originator regarding its use and purpose are ignored in the rush to build a justification for its use. So much criticism has been aimed at the framework over the course of years, yet still people recommend its use on forums, user groups and other social media for students and practitioners of evaluation.

One has to believe that, perhaps, vested interests are keeping the framework in place – that the name recognition and association with the concept keep it 'alive'. It took over forty years before the term 'Hoover' was partially replaced with that of 'Dyson' – perhaps it is the nature of change that we will be dealing with the concept for a long time to come. Meanwhile, until that happens, how many L&D practitioners will continue to struggle to 'do Kirkpatrick' rather than to begin to 'do evaluation'.

2.6 Section 3 – Training Evaluation – The Field of Study

As identified earlier, this phase is built on 'trawling' through a wide of literature to attempt to create a different insight from this review. In initially structuring the next stage of this chapter the ideas and frameworks discussed represent a range and cross section in three key areas:

1	Those who have adapted, or built upon the Kirkpatrick levels
2	Those who have moved outside the levels to create a different approach
3	Those who have innovated to begin to think differently about evaluation

Often, in literature reviews, selecting examples of the most prominent or widely referenced or cited authors is the method to discover the key ideas or themes. In attempting to remain effective and focused, whilst examining a wide body of literature to surface enlightenment, the

decision was taken to group the different 'modes' of evaluation thinkers. This was intended to create a 'logic' for the research as well as an aim that some fresh meanings may surface from taking this approach. As a starting point, the ideas of Landers and Callan (2012) helped to define potential groups as they proposed that there are three models of training evaluation:

<i>Process models</i>	These focus on the training design linking within the L&D process, such as learning objectives
<i>Hierarchical models</i>	These link together interrelated measurement outcomes such as competence etc.
<i>Meditational models</i>	These propose a causal effect between training and the organisation

Easterby Smith (1986) has also ascribed three elements to robust research and evaluation, namely, scientific, systems and naturalistic approaches and I decided to mesh the two concepts together to create a framework against which to evaluate and consider the role and impact of the various frameworks, tools and methods proposed and discussed in this review. The resultant framework is as follows:

<i>Process Focus School</i>	Where the primary focus is a process or systems approach to create an evaluation framework focused on generating process steps to create desired outcomes.
<i>Measurement Focus School</i>	Where the primary focus is to drive evaluation through the creation of new metrics or methods of measurement
<i>Outcomes Focus School</i>	Where the primary focus is the naturalistic and meditational generation of outcomes, within the wider societal or organisational context. Arguably, this approach links most closely with Decision Theory and may span, include or utilise elements of the two previous 'Schools'.

Within each of the sections will appear those frameworks that have built upon the Kirkpatrick approach, as well as those that owe their creation to a different source. Whilst a framework or evaluation concept may belong to a specific 'School', and some may well cross the boundaries, the motive in using this approach is to conjecture that a created grouping of concepts and themes may allow the opportunity of meaning to emerge, which could add value to the research project, as well as wider practice.

2.6.1 Group 1 – The Process Focus 'School'

Within the grouping I have proposed, Kirkpatrick's 'levels' were, arguably, the first example of the process approach. In fact it is the idea that the levels allow the perception of a process to be created may be its strength. What was missing, however, was a method of making the framework more robust and, whilst some thinkers in the field built an approach to merge the framework with measurement tool-kits particularly Phillips (1990), few successful examples were created.

The sheer number of scholars and practitioners that have remained within the paradigm is impressive and may reflect the belief at that time that the actual Kirkpatrick concept was worthy of development, perhaps because of its ability to appear as a process. Hence there is a constant development intended to 'fix' or 'improve' the framework. Whether these were consciously developed to add value to the Kirkpatrick idea is unknown in most cases; however, many refer to him and use his framework as a reference point for their own ideas and work.

It may also be useful to remember the various innovations in the workplace that will have contextually created different external 'pushes and pulls' upon thinkers in the evaluation field. These innovations particularly relate to the field of process optimisation, and include:

- The rise of Total Quality Management and Supply Chain Optimisation
- The development of Checkland's Soft Systems Methodology utilising Transaction Cost Economics from Coase (1937)
- Six Sigma, Kaizen and LEAN tools
- The rise and rise of the Balanced Scorecard Kaplan and Norton (1992), and their ubiquitous strategy maps
- Significant technological progress, including the rise of the Internet, Smartphones and Digital Media

Within the context of increasing change and technological advances in all areas of L&D, the evaluation field responded with a range of process-focused frameworks and tool-kits.

The CIRO framework Warr et al. (1970) consisted of levels, including Context, Input, Reaction and Outcome and had much in common with Kirkpatrick, in starting with the idea of levels. The idea of the framework was to situate the learning outcomes more closely to organisational requirements; although as Schmalenbach (2005) points out, "there is a lack of detail and prescription in how to undertake any of these four main elements". Although Passmore (2012) suggests that no internal metrics are created, the CIRO framework does share with Kirkpatrick the idea of utilising internal experts to validate the findings and create benchmarks.

The idea of internal experts to create answers is a frequently repeated theme and does raise a number of questions. For example: who are the experts? In what is their expertise? According to Warr et al. (1970), the answers at that time were no more than ways to spread the accountability for decisions, as most practitioners were working within the same paradigm.

The Stufflebeam CIPP model (1983) (Context, Input, Process, Product) arguably positioned as a 'systems model', sought to add value to the initial stages of the evaluation process, by improving the needs identification process. In defining a context in which the training was to operate, this created a method from which the internal logic of the framework could be used to evidence the training and learning. However, as Schmalenbach (2005) points out the fact that the Context stage points to a training solution that the model is somewhat biased in its approach.

Also, utilising control groups to drive the measurement of the outcomes restricts the concept in terms of ease of use and Bennett (1997), deemed that both the CIRO and CIPP frameworks are overly abstract and hard to implement in practice. Schmalenbach (2005) also identifies that the number of control groups and the elapsed time needed to make this large

scale process in evaluation effective, can be overwhelming. The idea of using control groups is also supported by Rae (1983) but with caveats, as he observes that social control groups do little actually to control the subjectivity of the result.

Scriven's (1996) Outcomes Focus process takes a contrary approach to the above, as it sets aside any specific process to suggest the novel approach where any process *"requires an external evaluator who is unaware of the learning programmes stated goals and objectives to determine the value and worth of that programme based on the outcomes or effects and the quality of those effects"*. This is an attractive idea for an external evaluator, but is vulnerable to the idea of bias and Schmalenbach (2005) contends that the principal idea does not lend itself to any metrics or tools of its own. Whilst the idea of bias and evaluator briefing may be difficult, it is arguably less problematical than relying on an internal verifier and the use of control groups.

Stokking (1988) proposed a process that blended together the Kirkpatrick levels with four additional perspectives or process steps. These include: customer satisfaction and societal contribution; linking in L&D needs assessment and planning; identified results and consequences, and availability and quality of resources. However, as Kaufman and Keller (1994) point out, there is too great an emphasis on the Kirkpatrick framework supplying the data to fuel the additional levels identified by Stokking and that, as Watkins et al. (1998) suggest, training alone will (or may) not achieve those outcomes, as the logic is corrupted by the tool-kit.

Brinkerhoff (1989) added extra levels to the Kirkpatrick paradigm, to consider formative and pre-learning aspects of evaluation. He then developed the idea of adding goal-setting processes to the design aspect, to ensure that there was an underlying sense of direction in the shape of the delivered learning. Whilst Holton and Naquin (2005) criticise the framework for being both formative and summative in nature, pointing out this is difficult to operate in practice, this framework is arguably the

first to set robust aspirations at the beginning of the design phase. The discussion about the nature of formative and summative evaluation really begins at this point in learning evaluation. Each has its proponents but those, including Brinkerhoff (1989), restrict their formative processes to a better understanding of learning objectives and some definition of stakeholder needs.

In considering different types of organisation, the assumption would be that the different drivers, contexts and stakeholder requirements would require a different form of evaluation, however, both Carpenter et al. (2016) (working in social work education), and Robinson et al. (2007) (inter-professional programmes), have heavily leant upon Kirkpatrick for their inspiration, producing very similar, expanded frameworks. Robinson however does stress the need to blend both formative and summative approaches; however, there is no tool-kit to help people deal with the process and this reinforces the view of Carpenter et al. (2016), that the whole idea remains 'somewhat mysterious'.

Robinson et al	Carpenter et al
Level 1 Reaction evaluation	Level 1 Reaction
Level 2 Learning evaluation - quality assurance index	Level 2a Modification of attitude/perception
Level 3 (Type A) Behaviour or skill application evaluation	Level 2b Acquisition of knowledge/skills
Level 3 (Type B) Evaluation of non-observable results	Level 3 Change in behaviour
Level 4 Bottom line impact	Level 4a Change in organisational practice
	Level 4b Benefits to service users

Whilst both are building on the Kirkpatrick framework and attempting to expand it, there is really no actual difference other than to situate it in their context through the use of appropriate language.

A number of other ideas are worthy of comment, Although most of them also simply build or adapt the Kirkpatrick model or attempt to use a

process approach to adapt the framework in itself. However, each incrementally adds to the body of knowledge within the paradigm and it at least helps those seeking genuinely to create something new to situate their concepts within proven methods.

Kearns (2005) added the idea of the baseline to all evaluation frameworks. Building on the TQM movement and Deming, the idea of training without a statistical starting evidence base became for him a point of criticism for many evaluation approaches. In this way, concepts such as the Bushnell (1990) IPO model (Inputs, Process, Outputs) lost validity as Robertson (2004) pointed out its inability to isolate a specific measure, as did the Hamlin (2002) 5 level model (which simply added a single variable into the Kirkpatrick framework) and the Guskey (2002) model which sought to make the Kirkpatrick framework more relevant to a university audience and the Kaufman & Keller Model (1994) which also changed the taxonomy to be more relevant to a manufacturing operation.

The Donovan and Townsend 'Nine Outcomes' (2004) attempted to help a L&D function organise its own process steps. The 9 outcomes consist of: Reaction to training; Satisfaction; Knowledge acquisition; Skills Improvement; Attitude shift; Behaviour change; Results; ROI; Psychological capital. The interesting idea is the effect the framework had on corporate image as it allowed the function to build Key Performance Indicators (KPI's) in each of the areas. This method expands the Kirkpatrick method into extra steps and can be seen as totally part of the paradigm. In its defence this is a trainer-practitioner process rather than an academic framework.

Some further examples of adaptations and advances in the process approach include: The Training Evaluation System from Fitz-enz (2001), which seeks to create a meta measurement system. Its steps are: Situation analysis – seeking an outcome; Intervention – training design and delivery; Impact – performance variables; Value – monetary worth on performance. This idea synergises some intriguing ideas around where value is created and how it is measured. Firmly part of the paradigm, this

is a pragmatic tool-kit for L&D people with a range of inputs and outcomes that help determine value. Heavily dependent on post-event fact finding, it is probably too reactive for the needs of today.

Pulley (1994) created the Responsive Evaluation Model to move evaluation to a different place, by being driven around the needs of the recipient. Its steps included: Identify the decision makers – to understand their needs; Identify the information needs – to generate the data to influence the decision; Collect data – qualitative and quantitative; Translate – into meaningful information; Involve and inform. The beginnings of the move from data and process to wider Decision Theory is a fundamental part of the process and serves to create a fresh perspective. This idea moves away from the concept of evaluation as an 'answer to an unasked question' to that more akin to decision theory and business intelligence. Whilst it uses the Kirkpatrick paradigm to generate the results it needs, it does start from a different point.

The Learning Effectiveness Measure Spitzer (1984, 2005) is a key framework in this group as it begins to move the thinking away from a process driven post-event measurement to some pre-event prediction. Its steps are: Predictive Measurement; Formative Measurement; Baseline Measurement; In-process measurement; Retrospective Measurement. It links advances in technology, strategy mapping and consideration of behaviour to align people and process activities to individual performance. The approach is designed to turn evaluation from a retrospective tool into a key Decision Theory approach by rephrasing the question that drives evaluation to be 'what should happen' rather than 'what has happened'. A number of the concepts in my own research will be driven around this developing concept in behaviour and process.

2.6.2 Conclusion

This group is characterised by a lack of real innovation with its primary focus in using process to build on the Kirkpatrick framework. Only Spitzer (1984) begins to use process in a more innovative way. However, there is now a straight line from Spitzer to Basareb (2007) and his 'Predictive Evaluation' approach in the move away from 'after the event evaluation' within the Process Focused School. The only restriction to innovation would be the 'handbrake' effect of the 'Measurement Focus School' that built its own credibility and tool-kits on the ideas of summative evaluation.

2.7 Group 2 – The Measurement Focus 'School'

The change in the context of the market over time affected by thinkers outside of the practise of evaluation shown earlier, influence this 'school' as they are more driven by data, metrics and evidence. This, more 'scientific' group, base their ideas more closely in the financial world and often, as time progresses, identify more with the concepts that 'big data' begin to offer. Some of the thinking from Human Capital Management and Human Cost Accounting Fitz-enz (2001) also begin to permeate the world of L&D.

The 2014 CIPD trend report identified that financial and analytical skills were particularly lacking in the L&D departments. It may also help to understand why, if the L&D function follows these approaches, they may find this school of thought so difficult and time consuming. However, it does raise the possibility that this 'School' is attractive to the function because the factual and analytical approach may have greater legitimacy with the wider organisation where those skills are in greater supply.

Some of the most significant scholars in this group include Flamholtz et al. (1969, 2003). They refined their approach of evaluating tangible and intangible assets into the 'Stochastic Rewards Valuation Model' that is built on the concepts within the field of Human Resource and Asset Accounting. The framework builds the economic value from a human

asset within the whole and expected 'service life' of an employee. The expected conditional value is then the expected impact from each person within their own projected personal 'service life'. The concept is influenced by 'economic value', which explains the present value of expected future services. By establishing progression and rewards for the individual across their service life, the stochastic process ('a natural system that operates in accordance with probabilistic laws' Flamholtz et al. (2003) can be determined by considering aspects such as hierarchical and reward progression.

In essence, this evaluation model calculates the hierarchical (succession) probabilities and rewards, with and without development, to establish the size of the uplift from learning. One criticism is that the concept only considers professional development qualifications and uses the finance community as exemplars - a function that has a more clearly defined career (and rewards) paths. Whether this process could work in less regimented functions is an area for further research. The framework is a triumph of the statistical notion that a strong logic allows for a mathematical formula to be created as proof of its own validity. For example, it contends that:

$$ERV = ECV \times P(R) \quad (2) \quad P(R) = 1 - P(T) \quad (3) \quad OCT = ECV - ERV$$

Where: *ERV*=expected realisable value *ECV* = expected conditional value
P(R) = probability of maintaining organisational membership *P(T)* = probability of turnover *OCT* = opportunity cost of turnover

Whilst being guilty of layering levels of subjectivity together in the pursuit of a tangible outcome, the framework is significant for a number of reasons:

- The establishment of strong logic and a link to Decision Theory
- The causal link between development and performance
- The creation of metrics other than ROI
- The idea of stoachism in a social science context of evaluation – this should not be undervalued as the creation of tool-kits that understand and measure the concept can be readily found in the

practise of risk management and is becomes a tangible bridge between summative and formative evaluation

However, it can be easily argued that it is too complex to be readily understood by a time poor and resource light practitioner base.

Whilst Flamholtz (2003) discussed evaluation of learning as a by-product of other activities, one person has created a dynastic organisation from his own concept and which could still become the rival to the primacy of Kirkpatrick in the field. This primary, and arguably best known framework in this group, is the Jack Phillips framework (1991, 1996, 1997, 2002), which is informally known as "Kirkpatrick Level 5" in the practitioner base. This approach eschews the traditional process approach and starts out with the need to prove a Return on Investment by linking together a range of costs and impacts, in order to create a financial impact figure. Described by Fitz-enz (1994) as 'the only evaluation approach that matters', the measurement criteria operates thus:

Sum of Total Benefits minus Total Costs divided by Total Costs and Multiplied by 100

The framework highlights the need to become rigorous in the area of cost identification by allocating direct, indirect and opportunity costs against both tangible and intangible benefits. It is structurally a summative process but it does begin to highlight the need for some specific learning goals and impacts to be created as an initial aiming point.

Whilst McGovern et al. (2001) are enthusiastic proponents of the framework in the coaching world, they state that, at its heart, are a series of estimates and assumptions that Passmore et al. (2012) state are not credible in what is meant to be a more robust, scientific approach. Considered by many evaluators as the 'gold standard' of financial impact evaluation, it is extremely expensive to operate. It is also seen as the process most likely to be replicated in other fields of social impact evaluation as a strong summative evaluation process. However,

to any financial accountant, this is no more than a simple finance equation applied to evaluation.

The process was used by Wills et al. (1996), to establish the Return On Investment ('ROI') of a longitudinal study of Action Learning MBA delegates where over 5000 managers were developed at a cost of around £3million. The managers spent approximately ten million hours in development across the life of the programme. A huge range of financial and non financial impacts were identified, leading to the claim that the 5000 delegates triggered over £100million of impact with ROI of nearly half a billion pounds. Whilst the figures are large, Phillips suggests that the rigour in the process allows for the collection of tangible data from which strong and legitimate conclusions can be drawn, even when highly subjective, particularly when using skilled resources to generate and interpret the data.

Burkett (2005) built on the Phillips method with some particularly useful examples in helping delegates ascribe value to intangible effects of learning. However, whilst using the process of estimations, weighting and confidence intervals is statistically satisfying, it does reinforce subjectivity – even if supported by a strong sense of logic. However, her starting point is to take a cost saving approach to the learning process and uses ROI metrics to demonstrate savings. This could be problematical as the approach may affect the true value of the ROI number, as the simplest solution to increase ROI is to reduce the cost element of the equation, rather than to seek enhanced value.

The principal metric at the heart of the Phillips method is ROI. However, the phrase 'Return on Expectations' (ROE) has also appeared in recent years, based on the idea of a more thorough examination pre-course of the various needs and outcomes required from the programme. Anderson (2007) in her research for the CIPD is an enthusiastic proponent of the 'metric'. However, Phillips & Phillips (2011) are very clear that this is a term that generates confusion, adds little value and is 'no more than a means of describing of Kirkpatrick Level 1'. The Kirkpatrick foundation in

their current marketing literature also agree with this description even though they have rebadged the Four Levels as a ROE process.

The debate whether ROI is the correct measure at all for the measurement of human endeavour rages across L&D, Luthans et al (2006), Phillips (2007), Fitz-enz (2001), Kirkpatrick (1994); without arriving to a conclusion. Kearns and Po (2012) posit that any measurement of 'soft-skills' attempting to generate ROI is intrusive and can actually destroy the intended value of a programme. Goldsmith and Sarno (2009) also produce a list of ten reasons why ROI is a poor measure for training with the most salient reasons being:

- the lack of meaning over time
- the over focus on cost
- the subjectivity of data collection and misattribution
- the under-reporting of risk and opportunity
- not addressing the variables that matter or that are valued

Botchkarev and Andru (2011) have determined that ROI is less useful as a concept in practice than in theory. This is partly, as they contend, because L&D practitioners often use the terminology without actually understanding what ROI is or how to conduct a ROI analysis. They conclude that ROI was originally designed in finance and meant to be a one-time measure of a capital project and training is not that type of process. They also conclude that ROI can overly focus the L&D function on short-term returns at the expense of strategic KPIs. They conclude that: *stretching to improve ROI is sub-optimal in terms of the firm's goals (profitability, cash flow or shareholder value) and will typically cause expenditure to be lower than the firm may want it to be.*

However, the findings earlier in this review show that the L&D function operates within the Kirkpatrick paradigm and this approach is, in effect, Level 5, so ROI (however interpreted) continues to have primacy in the metrics field, even if only in the language. Seen as an adjunct to the Kirkpatrick paradigm, it is seen as evidence and gains legitimacy by

having become so associated with the core paradigm, through a process of mutual reinforcement.

If Kirkpatrick (ROE) and Phillips (ROI) represent the primary school of thought in evaluation metrics, then the work of Cascio and Boudreau (2011) has formulated and influenced significant number of the 'alternative thinkers' by providing concepts and tool-kits based on alternative evaluation metrics as their world-view is firmly rooted within the field of Utility Theory. Utility Theory had existed as a concept for many years, for example in decision-based utility models, arguably originating in the work of Brogden (1949) and Cronbach & Gleser (1965). Arguably, Utility Theory had been a solution looking for a problem until seized upon by the Human Capital School who adopted it to use as the basis of measurement for showing the impact of 'people on the bottom line', as well as the HR and L&D processes that drove those results. Inevitably, it began to be used within the evaluation sphere as well.

Cascio (1976) and Boudreau (1983) proposed that Utility Theory and Capital Budgeting Theory be integrated in order to create an understandable tool-kit to generate financial returns from development. Those 'easily understood' tools included payback, break even analysis, decision trees, sensitivity analysis, discounted cashflows etc. They positioned the concepts as a summative process which then had inadvertently added value to two other concepts: the Phillips method by linking robust financial measures to ROI as well as increasing the potential legitimacy of formative evaluation by linking the it to robust financial modelling methods.

In addition to this, Cascio (1987) developed the idea of Meta-analysis to be able to determine standard deviations using large data to help show gained outcomes. These were initially applied to selection methods as reported by Schmidt et al (1982) and later for more training interventions as shown by Guzzo, Jette & Katzell (1985) and Cascio (1985). Cascio grasped that this estimating process could be used to predict potential returns but had insufficient data or computing power to be able to

operate a robust process. This idea of being able to calculate the financial benefit of a standard deviation would seed an idea for evaluation as they developed new Utility Theory constructs and tool-kits.

Developments in the concept of Utility Theory continued to attract attention in evaluation due to the work of (for example), Weinrich (2001), who defines multivariate utility analysis as having three elements:

- Group utility - inspired by utilitarian moral theory
- Intrinsic utility - inspired by value theory which depends on goals
- Expected Utility - inspired by probability theory and goals

In this way a utility can be positive or negative and the utility can be divided by the probability to show efficiency and the value of the impact of the utility. This allows for the potential of further research within the HR and L&D fields to build tool-kits as Weinrich (2001) states that the ability to build 'value' around positive and negative utility can be applied to a range of 'soft' or intangible methods.

However, there are many critics of the whole idea of Utility Theory and the linkages between the theories and the tool-kits. Latham and Whyte (1994) built their research on the ideas of Schmidt et al. (1982) who found that practitioners are sceptical of forecasts from utility analysis, primarily because of the difficulty of creating credible dollar attributions from standard deviations in behaviour. In their sample of 143 Managers, Latham and Whyte (1994) found robust statistical evidence of performance improvement but that the results were not trusted because of the implied simplicity and overstatement of performance gains relative to the intuition of the management group. In fact, managers trusted the results less because of the use of Utility Theory. This result was further reinforced by Whyte and Latham (1997) when the experiment was repeated and an expert in Utility Theory attempted (and failed) to build increased credibility with the research base.

One of the features of Utility Theory and the associated analytic toolset is their simplicity and this has bred problems in the research base. Sturman (2000) found that the overly simplistic nature of Utility analysis did not take into account programme specific factors and thus the outputs lost validity even though Cabrera and Cabrera (1999) see a more strategic use of Utility Theory when integrated with the Balanced Scorecard concepts of Kaplan and Norton (1992, 1996, 2010). It is also worth reflecting that Mintzberg (1975) suggested that managers pursue a project on the basis of who supports it rather than on the basis of project quality and that Utility Theory was less well supported by the 'better known' thinkers of the day and he conjectured that this would continue to affect its credibility.

In 2011, Cascio and Boudreau collaborated to create their joint approach to HR and L&D metrics that linked many of the Utility Theory and financial tools together. They also demonstrated that their evaluation framework showed a correlation between performance uplift and standard deviations. This build upon their previous research where Cascio (1982) and Boudreau (1983) had established that training had a diminishing impact over time and needed to construct a metric to address the reduction in impact. The idea would be more fully developed by Baddeley (1999) in an associated field of thought, building on the Ebbinghaus (1885) idea of the 'forgetting curve'.

Whilst initially using a productivity indicator as a means of measurement behind the Human Capital theory of Engagement are now beginning to see Utility Theory as a unit of justification for engagement programmes. (This is an area being used within my own research framework to build an evaluation strategy and one, that causes the most consternation in the minds of the L&D function with whom I work).

In creating tool-kits such as these, Cascio and Boudreau (2011) laid the foundations for a more financially robust approach to both formative and summative evaluation such as the TDRp (Talent Development Reporting Principles) developed by Robert Zeinstra in the Toyota Academy which

are used in Toyota as a basket of measures which can be used as part of a scoreboard as well as other measures such as 'EVA' (Economic Value Added) that are used in the field of Human Capital Accounting. However EVA does not have a foothold in evaluation at this time and can be safely ignored in this document.

Akrofi et al. (2011) have also innovated within this group of thinkers, although admittedly focusing their approach on specific measures of executive or management development. Their work is built on the concept that leadership development is much more about the whole person than a specific role, and this can create problems for standard evaluation approaches. They cite the need for a move towards increased evidence of the following activities to demonstrate task competence of leaders in a rapidly evolving environment as evidenced through "informal activities such as coaching, Jones et al. (2007); Thach (2002); action learning, Seppanen-Jarvela (2005); self-development processes, Baruch & Hunt (2003); mentoring and other peer-related learning activities, Cullen et al. (2004)".

The learning development measure they have created is defined by four main dimensions: strategy, self-direction, experience and participation and builds on measures of human capacity to link those together with leadership competence. A high level of statistical computation unusual in evaluation literature is used including Confirmatory Factor Analysis and 'Goodness-to-fit' concepts (as well as peer review), providing evidence of the robustness of this framework. Their conclusion that this type of development produces increased or enhanced leadership competence is reinforced by a comparative summative Phillips' evaluation of leadership effectiveness used within the same study.

Another measure that has gained increasing acceptance in L&D is the use of the Net Promoter Score ('NPS') a measure derived from the world of marketing as it posits a link between buyer satisfaction and intention to recommend or repurchase. A number of training companies (including 'Metrics that Matter' one of the largest and most influential evaluation

software organisations) have championed the concept and it has gained traction because of its simplicity and application (with varying degrees of credibility) in other areas of the organisation.

Fred Reichheld launched the NPS concept in 2003 as a marketing tool-kit, where customer satisfaction theory had been looking for a simple measure to demonstrate its validity. The irony of this position is highlighted when thinking about the world of evaluation! Marketers were quick to embrace the score that divided respondents into those who could be categorised quickly into 'detractors' and 'promoters'. The score could also be used as part of a further series of activities to build plans and activities for subsequent process improvement. This use reflects its potential use as a Decision Theory tool. The NPS concept has been adapted by the supporters of the 'Word of Mouth Index' that creates a more robust adaption of the NPS concept.

Hanson (2011) identifies that both supporters and detractors of the idea argue back and forth about a range of problems including:

- Its supposed links from customer satisfaction to buyer behaviour
- Its ordinal scale and inequality of measurement
- Its predictive validity
- Its simplicity

However, evaluators have also begun to link this measure into training, arguing:

- A causal link between training satisfaction and post-course behaviour, even though this is a subject of much dispute as reflected in the work of Noe and Schmidt (1986)
- A strong scale and clear outcomes
- Its predictive validity (not proven)
- Its simplicity

As a metric that has gained traction very quickly, the risk is that this could be a measure that L&D could embrace equally enthusiastically and with as little enlightenment as other metrics and frameworks. Metrics that Matter (on their website) have suggested that the NPS score is a suitable measure of Level One of Kirkpatrick. The idea of two unstoppable and simple ideas coming together in this way may be the quick and easy evaluation solution that L&D seeks, even though arguably, that multiplying something that doesn't work with something else that doesn't work should not equal something that does!

2.7.1 Conclusion

This school of thought has operated independently from the Process school, however, one has used the other to justify or build different approaches.

In reality, when thinking about the time period over which this document spreads, from the creation of the Kirkpatrick framework (1954), then there is really a paucity of innovation in this area. Each new improvement or innovation has been under-respected unless either linked to a solution for the deficiencies of Kirkpatrick or operating as an ally or to enhance the existing paradigm.

Other approaches not directly linked to Kirkpatrick risk being ignored or forgotten. In effect, still the most credible alternative approach is the 'rebadging' of a simple financial idea (ROI). One hopes that the rise of 'big data' and some of the innovations in measurement mentioned later in this document could provide the fuel for some further innovation in this area of evaluation.

2.8 Group 3 – The Outcomes Focus 'School'

The idea of this section was to examine those thinkers who had created both measurement and process approaches or those who had used a more holistic starting point, rather than just having an evaluation focus.

Many of these approaches appear to use a conscious or unconscious informal Decision Theory approach to help posit and resolve questions.

One of the principal concepts in this section (and one that has heavily influenced my own thinking), was first proposed by Kraiger, Ford, and Salas (1993). Building on the taxonomies of Bloom et al (1956) they moved away from the idea of learning as evaluated only by changes in verbal knowledge. They then proposed that learning outcomes would be based around three areas: cognitive, skill-based and affective outcomes. Each area was then supported by a range of constructs to help measure and understand both the concepts and the outputs. The evaluation method is then reinforced with a pre and post-process described as 'probed protocol analysis'.

Area	Protocol
<i>Cognitive Outcomes</i>	3 Indicators: Verbal knowledge, Knowledge Organisation and Cognitive Strategies Process: Organise and process knowledge – build procedural knowledge through practice – apply to real life situations
<i>Skill-based outcomes</i>	2 Indicators: Technical and Motor skills Process: Replicate behaviours through copying & practice to become more fluid – then apply in the real world
<i>Affect-based Outcomes</i>	3 Indicators: Motivation, Self efficacy and Goal-setting Process: Use with 'compilation' to generate change

Whilst criticism exists that some of the concepts lean too heavily on Kraiger's own research, Alvarez et al. (2004) found legitimacy in many of the core assumptions. Additionally in 2003, Patterson and Holey investigated work by Ford Learning Academy using the concept and showed the method to be both robust and useable by practitioners, even in the area of 'soft skill' development. Possibly the most realistic criticism from Griffin (2011) is the nature of change and evidence that some of the core ideas may have been superseded by new learning about effect and motivation as discovered by David Rock (2008) in the field of Neuroleadership.

Whether Mayne's (2008) idea of Contribution Analysis should sit in this document at all is a moot point, as it is primarily intended (or used) in the social impact space, rather than the evaluation of learning. However, it was recently used in a large NGO as an overlap into training evaluation and it has features that make it interesting to this body of knowledge and therefore worthy of a brief review as it aims at identifying causality from a range of variables to an output and therefore has legitimacy as a L&D evaluation method.

The 6 steps include a definition and a tool-kit of questions to ask as an evaluator. These include:

- 1 Set out the attribution problem to be addressed
- 2 Develop a theory of change/logic model
- 3 Populate the model with existing data & evidence
- 4 Assemble and assess the 'performance story'
- 5 Seek out additional evidence
- 6 Review the performance story by repeating 3-4 until satisfied

This broadly summative approach has all the weakness of other subjective ideas at its core but has a strong logic and is useful in Decision Theory to find out 'what worked, or not'. However, whilst it has much in common with the ideas of Rational Evaluation, Pawson and Tilley (1997); it is acknowledged as being more useful in organisational cultures as it acts as a 'whole system' improvement approach using evaluation processes within it to make the decision process effective. The inherent complexity in this approach means that it is unlikely that a practitioner in L&D evaluation would use this framework outside of a whole system approach.

However a more accepted and pragmatic approach is from a contributor to the field whose work has been previously mentioned. Brinkerhoff (2003) developed one of the most useful evaluation frameworks in use in the world of evaluation, which focuses its attention on Decision Theory

and continuous improvement. The Success Case Model (SCM) arguably owes much more to social research methods than Kirkpatrick and, whilst being completely summative, it aims to use a different approach to gain insight from which decisions can be made. The framework uses a sampling method to identify the highest performing delegate to show the possibility for impacts from the programme. The process asks core questions to determine what is happening post-learning, what results are generated and the value of those results and how could the initiative be improved.

Whilst Casey (2006) identifies a level of subjectivity in both the quality of delegate responses and the analysis of the data, the SCM is recognised as a valuable method for both the practice of the external and internal evaluator. However, one of the challenges for the concept is that it is rarely used on its own, being subsumed within the Phillips' summative process and, as such, loses some of its ability to drive insight.

Ed Holton also appears in this group with a different concept from that outlined earlier and shows the progression of his thinking. In 1996 he created a new framework involving three levels (learning, individual performance and organisation) broken into primary and secondary factors, to create outcome measures. From these, a process could be created to drive these outcomes. However, in 2005 he was forced to amend the concept as Kirwan and Birchall (2006) criticised the concept for its lack of feedback loops and interaction. However the concept itself is interesting as one of the first attempts to link evaluation and Decision Theory. He then refined this 'Results model' to separate properly the processes of assessment and evaluation within the new framework, giving clarity to evaluation outcomes including: System and financial outcomes; Learning knowledge and expertise outcomes; Participant and stakeholder outcomes. Holton's contribution as an innovator in the evaluation field through a number of credible tools and approaches is recognised by a range of practitioner groups.

Little (2004) created a pragmatic starting point to examine stakeholder outcomes and argues that his meta framework for evaluation should be used to generate ROI by using best practice strategies, including:

- Alignment to organisational objectives
- Self-paced and learner focus
- Designed to embed into the workplace
- Granular, bite-sized chunks
- Delivered via a multi-media approach

Whilst this concept builds on the work of Geertshuis et al. (2002), this approach favours a specific type of learning intervention and may not suit the needs of learning that mostly benefits the organisation – particularly because of the need for speed of transfer and links to a wider goal. Also, the tools included consist of the usual summative processes, (interviews questionnaires, focus groups etc.) as well as methods particularly geared around e-Learning process outcomes, which does suggest the use of this approach more for that learning process.

A more interesting meta framework is that produced by Dessinger-Moseley (2006). The Full-Scope Evaluation Model (FSEM) blends formative, summative, confirmative *and* meta evaluation processes. This idea is built on and heavily influenced by the idea of 'comparing results with intentions' a key part of the methods from the work of Kaufman et al. (1995) in creating a rival approach to the Kirkpatrick framework. The four stages of the FSEM process are built around a number of associated activities:

Stage	Activities
<i>Formative Evaluation</i>	Analysis of cause, current performance levels and the appropriate selection of design responses
<i>Summative Evaluation</i>	Reviews of reaction and competence
<i>Confirmative Evaluation</i>	Understanding the values of longer term impacts (including ROI if required), the building and reinforcement of learning (learning transfer)

<i>Meta Evaluation</i>	Ensuring the evaluation itself is of appropriate quality and has appropriate insight
------------------------	--

The authors themselves acknowledge the difficulties inherent in the process and the need to engage the attention of the organisation for a long period of time to prove impact. However, the blending of the different processes within the framework is a genuinely interesting approach and one that has a great deal of potential for practitioners in being able immediately to grasp the concept of the clear stages of evaluation. However, the lack of practical tool-kits is a problem for ease of adoption.

It is worth pausing to consider the timeline for this group as the pace of change and the number of ideas and concepts appearing within this group or 'school of thought' represents more imagination and innovation than those in the other 'schools'. It is also worth noting that the economic picture from 2007 changed considerably and the need for cost control and value perception became increasingly important as recessionary forces became encompassing and pervasive.

In this context, Valerie Anderson (2007) created a new evaluation model as part of work conducted for the CIPD. As this is the professional body for the function, it is worth including an overview here. In the research she identified the need to view needs through a series of different lenses to ensure that clarity of concept and approach was created. She also concentrated on the needs of the management population and organisation rather than the needs of the learners. She developed a hierarchy of outcomes based on four areas of summative and formative evaluation that, in my view, was inspired at its time of publication:

- Return On Investment – To what extent is the learning contributing to pre-agreed learning objectives
- Return on Expectation – To what extent have the expectations of the stakeholders been met?

- Benchmark and Capacity measures – To what extent does the function operate effectively?
- Learning function – To what extent does learning function deliver?

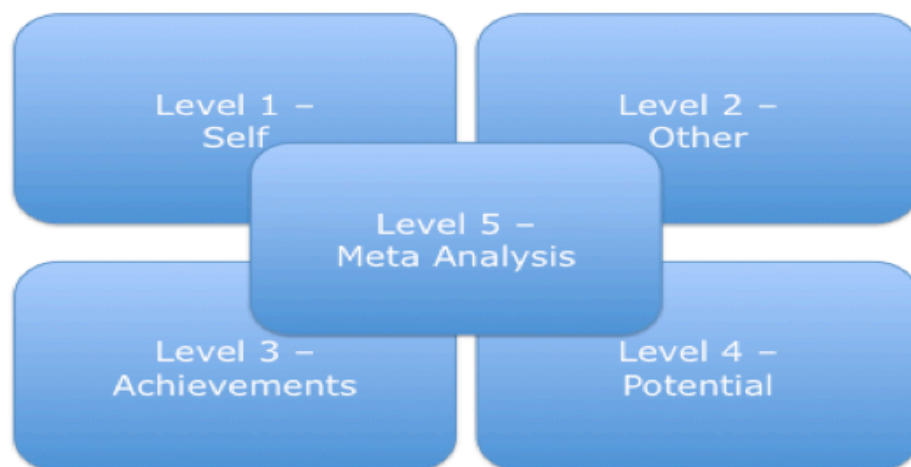
Arguably, although being influenced by the Kaplan and Norton (2002) Balanced Scorecard, the ideas were interesting, however, it can be argued that they have failed to take hold and appear to form no part of the syllabus trained by the CIPD around evaluation in their current qualifications.

Many of those academics that have criticised the work of others have failed to provide a convincing solution themselves. Of those, Richard Griffin may have done more than most to create a satisfying solution to the idea of an integrative approach. Griffin (2011) has generated a meta framework, bringing together many of the most obvious elements of processes, with suggested tools at each phase and has linked together both formative and summative elements with stakeholder needs. The framework is shown in Appendix 2. Whilst it does build upon the Kirkpatrick framework, it more closely mirrors the reality of the L&D operation. This work is built upon his belief that evaluation is an emergent field and it attempts to bring order from the various themes and competing ideologies. It also builds on the contention of Chiaburu and Lindsay (2008) that the correct learning triggers must be in place to drive the appropriate levels of engagement and evaluation.

Griffin also stipulates that proper learning transfer must take place, as the application of the learning is key to make the training worth the investment. This can be a problem as learning transfer is often not part of the remit of the L&D function, being more organisationally situated in the role of the line manager. In fact, Burke and Hutchings (2007) identify a range of contingent factors to this learning transfer process and the way this framework gives embedding due significance may be the difference for its successful application by integrating those factors.

A more holistic approach to evaluation depends somewhat on the definition of stakeholder needs and Guerri and Vinate (2011, 2011) created a holistic set of principles based on the work of Rossi et al. (1999, 2004). These included the need for Efficacy, Efficiency, Accessibility, Image impacts, Transferability, Innovation, and Synergy within the evaluation process, in order to be able to create the required stakeholder outcomes. Whilst there are no measures or tool-kits proposed, their research did highlight the lack of focus on specific stakeholder needs, perhaps highlighting how much of L&D process exists outside of stakeholder awareness or relevance.

Passmore (2012), noted for his criticism elsewhere, created the SOAP-M framework, which blends together five levels of analysis: Self, Others, Achievements, Potential, and Meta Analysis to give an integrated outcome for a programme.



Each Element is linked to a range of tools to help create an integrated framework, thus:

Element	Measures etc.
<i>Self</i>	Individual Psychometrics such as for Emotional Intelligence (EI) or resilience, personality etc.
<i>Other</i>	Behaviour

<i>Achievements</i>	Task Achievement through SMART goals – or target achievement
<i>Potential</i>	Assessments of potential – perhaps through 9 box succession etc. as well as individual psychometrics
<i>Meta Analysis</i>	Wider, organisation-wide data analysis

Whilst the framework contains many of the problems that Passmore has criticised others for, including a lack of proper rigour in the tool-kits and an over-reliance of subjective internal processes, it is refreshing to see the linking together of more strategic measurement criteria with individual potential – arguably the first time this idea has been proposed. One of the issues is also the over-reliance on ‘weaker’ conceptual psychometrics – the rise of neuroscience and models of personality may throw doubt on some of the traditional approaches using simple factor analysis in the questionnaires he proposes.

Learning Transfer is an associated ‘field of thought’, which overlaps with evaluation, as it co-exists as both the means to carry out the evaluation and the process to ensure that learning transfers from concept (or procedural knowledge), into utility (or performance) in the workplace. There is a huge field of research in this area but much of it comes to fruition in the work of Holton et al. (2000) and the development of the Learning Transfer System Inventory (LTSI) instrument. Holton contends that designing the ability to transfer is the same as designing good training but that the measurement and support of the learning actually happens at different stages, extending beyond the training room. This idea is also supported by ‘Spaced’ and ‘Sticky’ Learning concepts - Holton (1998) - which use post-course blended reinforcement of key messages.

Ford et al (1998) reviewed twenty empirical papers with a blend of quasi-experimental and experimental projects, to help ground the tools and techniques of applied transfer in applied research. In the end, four broad areas were established as being conducive to good transfer. Learner Readiness, Management Support, Team Engagement and Embedding Processes. This was further refined by Fecteau et al. (2012) into support needed from the four parties of top management, supervisors, peers, and

subordinates, in addition to the learners themselves and, in particular, their pre-training motivation. Whilst Blume et al. (2012) differentiated transfer needs and processes between 'closed' (e.g. IT skills) and 'open' (e.g. leadership) types of training, they were broadly agreed that transfer was a key process for the focus of the L&D department, particularly in being able to evaluate impacts.

Roullier et al. (2006) also stressed the need for an appropriate organisational climate to be created in order to produce required returns. Their study of retail franchisees indicated the correlation between more effective transfer and climate. Their conclusion that climate was more important than the volume or nature of learning within the course is a major factor when constructing an evaluation process. It is also particularly important when also considering Little's (2014) claim that 50-80 per cent of activities on a programme are defined as 'scrap learning' and add little to no transfer value. This approach allows learning to be better constructed and reduced to more important elements (learning points), to maintain learner motivation.

This research highlights the need for a continuing dialogue between the Evaluation and Learning Transfer schools of thought to push forward this mutually beneficial creation of concepts and tool-kits.

2.8.1 *A challenge in the literature*

It was late in 2015, and some years into the development of my new evaluation framework (shown in the research phase later in this document) that, whilst in a LinkedIn group I stumbled upon Dave Basareb and his work 'Predictive Evaluation' – first published in 2011. I recognised an initial overlap in our ideas without having any prior sight of any papers or articles or any materials he had written. He proposed that a simple process of formative evaluation was all that was required in order to build a strong and believable evaluation process. He had also moved away from the process of 'Reaction' to 'Intention', thus mirroring

my own work and research. Whilst this formative approach has many critics, I find it hard to disagree with many of his ideas, as they so closely reflect my own. His process also built on the work of Guba & Lincoln (1981) and placed huge emphasis on collaboration and negotiation among all the internal stakeholders to 'socially construct' an approach and a mutually agreed set of outcomes.

Within the text of the book are few academic references, however; as it is aimed at the practitioner, this is understandable. To this extent I have shown below, for the sake of clarity, the primary similarities and differences to clarify our approaches.

Basareb	Thackeray
Creation of Training Strategy Based on KPIs	Creation of Training Strategy based on 4 'value quadrants' Valuation of the L&D budget
Building of value proposition based on internal Delphic process	Building of value proposition based on external Delphic process
Measuring Intention as a core KPI	Measuring Intention as a core KPI but based on the work of Anscombe within Action Theory
Seeing goal theory as the key driver in generating impacts	Seeing intention theory as the key driver in generating impacts by linking to Learning Transfer
Linking behaviour to results through a dashboard of KPIs	Linking process change to results through a dashboard of KPIs
Return on Expectation	Return on Targets

Many of the key criticisms that could be levelled at the Basareb framework could also be aimed at my own research output. However, in creating a user-friendly approach and tool-kit we have both independently sought to deliver an evaluation solution that is as simple as Kirkpatrick and as robust as Brinkerhoff. Interestingly, his ideas are endorsed by Donald Kirkpatrick in the introduction as a legitimate way of calculating Return on Expectation. At this stage, there is little academic critique of the Basareb ideas, as they appear in the practitioner press

rather than in the academic literature and, whilst it is disconcerting to have discovered something so similar, it is an encouraging feature of research to have different people making similar discoveries.

2.8.2 *Review of Methods and Conclusions from the Evaluation Literature*

It is appropriate in reviewing the literature to comment on both the frameworks and metrics that became outputs from research as well as some of the methodological considerations that shaped those outcomes. This would help serve as a guide or assessment point for the choices in my own research across the life of the project.

It is striking to see that the schools' identified in this literature review tended to follow similar research processes within the grouping. Those in the Process school tended to follow a broadly inductive process carrying out either qualitative or quantitative research and those in the measurement school tended to reflect a more deductive approach and substantially greater use of quantitative methods and hypothesis building. Interestingly, the hypotheses seeking a metric or proof of the efficacy of a metric also used substantially more financially robust methods rather than simply relying on social proof. Those in the 'Outcomes Focus School' unsurprisingly used a blend of inductive and deductive methods but with a greater use of qualitative data.

Kirkpatrick (1954) in his original treatise used an inductive process in his PhD research using a case study approach and interviews to assess the levels of learning that became the foundation of his overall approach. Those most influenced by his research also tended to follow a similar approach, for example, the CIRO framework (Warr et al. 1970) was built on interviews and focus groups to determine the adaption of the framework.

Interestingly, the use of control groups became more commonplace with the work of Stufflebeam (1983) and Stocking (1988). The use of

comparative data using a large sample of respondents allowed them to propose amendments to the Kirkpatrick framework. Criticism of the method was presented by Rae (1983) and Bennet (1997) although their objection was as much about the use of Kirkpatrick as a starting point as the use of control groups per se. Scriven (1996) represented an interesting conjectural sub theme through a more innovative methodology by utilising a thought experiment to propose his idea. This builds on many of the ideas being presented within Kaisei and Total Quality Management that had also shown the need for third party validation of results from an independent idea generator and evaluator, preferring to separate the various parts of the process.

Kearns (2005) also used the methodological toolkits espoused by the management processes in Total Quality Management rather than those of social research. He preferred to use a highly statistical approach, reviewing the actual evaluation process and improving the process through a statistical deviance review. Other practitioners also used a more conjectural and statistical approach including Donovan and Townsend (2004) to create their framework. The move away from a more traditional social research method was aided in part by the advances in computing power, that helped to statistically model the changes in process rather than relying on more traditional research methods. The Measurement focus school were broadly able to use a more deductive research approach as they were testing the usefulness of the application of the metric rather than surfacing meaning. Phillips (1982) was at the forefront of this move applying a standard financial metric to evaluation. His initial deductive approaches built on case studies of existing research to create outcome in the use of the metric. Phillips continued with this approach over another three iterations of this method (196, 1995, 1996) to collect a huge range of data from evaluation projects completed with clients and through continuing deductive academic research. He was arguably well placed to use the approach of the Law of Large Numbers which is the bedrock of the approach in this paper.

The financial metric of ROI is often derived from statistical review of learning outcomes collected through summative evaluation. Wills et al (1996), Fit-enz (2000) and Burkett, (2005) all were proponents of this approach and built on the Philips hypothesis to further refine the use of the metric as being useful. The use of Case Study is particularly striking throughout this group with little or no use of action research. Kearns and Po (2012) comment that this is because the nature of statistical, qualitative data collection and analysis and the need for a suitably robust contextual framework.

This methodological approach of analyzing data from Case Studies using summative data collection and deductive analyses was used by Cascio (1976) and Boudreau (1983) when they were proposing and testing the initial concepts of the use of Utility Theory as an evaluation metric to rival ROI. Whilst the methodological stance was similar to Phillips and others in the field, the lack of very large data sets and the computing power needed to analyse them led to the dismissal of both the statistical analysis and their conclusions. Latham and White (1994) were instrumental in the initial demise of Utility Theory because of their simple argument (published in the trade press) against having used a small sample deductive review of opinions from interviews. It is interesting how in this area of robust financials, that a qualitative approach was able to scupper the fresh idea proposed by a more robust deductive method simply because the prevailing paradigm was being reinforced by the work of Phillips.

The rise of NPS as a credible measure has been the subject of intense scrutiny in the Marketing press with a range of research processes both generating approbation and disapproval. A simple deductive measurement is used in summative evaluation of L&D programmes but Hansen (2011) in his qualitative work in validating the use of NPS in training evaluation is still dubious of its applicability in this field of practice.

The final, more holistic school of thought, blended and built upon a range of methods but often deployed Case Studies as a means of collecting both qualitative and quantitative data around which they surfaced meaning or proved hypotheses. Brinkerhoff, (1989 and 2003) with a background in social research initially focused on the traditional inductive approach, but later moved to an overt deductive process when validating the hypothesis behind the Success Case Method (SCM). His ability to operate the SCM across a range of case study clients allowed a strong statistical justification for the use of the method.

Holton and the Learning Transfer method operated both Action Learning and Case studies to assess the efficacy of both transfer of learning and the Meta evaluation data that was surfaced. In over twenty five projects over half of them were able to roll through at least four iterations of data collection to allow for proof of concept as well as to satisfy some of the conditions required by action research. The requirement to be embedded long enough within a corporate environment does mitigate against the use of Action Learning as pointed out by Kraiger et al (1993) and often results in the use of Case Study to generate results that are both 'relevant and pragmatic'.

Griffin (2011) along with Dessinger-Mosely (2006) built phases of research methods upon each other as they blended independently operated projects into a more complex framework. This ability to blend together a range of psychological, social and statistical methods allowed for a more holistic output from their research.

Ford and Weissbein (1997) adopted a Meta approach to develop their framework based on a range of empirical papers from the wider literature. This work became an integral part of the Learning Transfer approach and added value to the more traditional approach operated by Holton outlined earlier.

A subject such as evaluation that is itself part of a social research movement is often constrained by the paradigm and assumptions that

surround it. Kirkpatrick and Phillips have contrasting approaches but, taken together, offer a credible and robust approach. The work of the Utility Theorists may have been rejected because of their decisions to separate themselves from the prevailing paradigm but Spitzer (1992) and many of the other Metric 'school' were able to gain traction within the field of practice without that requirement.

What is clear from the research is that, as evaluation is situated within an internal organizational process then the use of a Case Study is an effective and pragmatic means to begin to find, collect, organise and analyse data.

The enthusiasm I have in the range of ideas and innovative processes in this group is reinforced by the fact that some of these approaches are in place in some organisations despite their relatively late development. Kraiger (1993) is seen as a legitimate force in the area and has transformed thinking in summative evaluation. Brinkerhoff and the Success Case Method is regularly discussed in forums and implemented widely by evaluators.

Griffin (2011) has some interesting ideas, but disappointingly appears to have moved his area of research away into the world of coaching. Perhaps this reflects the biggest challenge that, in order to develop a commercially viable practice, few people 'make it' in evaluation, other than those who continue to champion or propagate the core paradigm.

Given the nature of L&D and its ability to engage with the latest ideas in learning, it is by nature depressing that evaluation seems still to be so low down the list of priorities. The only glimmer of light, perhaps, is the moves and processes within Learning Transfer. The LTSI (Learning Transfer process evaluation tool-kit) actually measures impacts using Utility Theory, a key evaluation idea, but it is positioned as a learning transfer tool. In reflecting on the approaches above, it becomes apparent to me that a number of them appear to have driven the thinking in the field:

- Thinking about evaluation as a process (or hierarchy) in itself with 'tool-kits' to make it work
- Thinking about evaluation as a collection of measurements (supported by tool-kits; both shared with, or separated from, the process approach)
- Thinking about evaluation as part of a Decision Theory approach
- Thinking about evaluation as an outcomes reporting method, again with tool-kits to support this and shared with the areas above

One of the issues is that unpicking the differences between each of these overlapping and competing ideologies is still a challenge for the L&D practitioner. In building new approaches or spawning imitators, the ability to create a simple, new paradigm has, therefore, been sabotaged at the most basic level from being articulated as something which is as 'practical' and 'straightforward' as Kirkpatrick – even though Kirkpatrick 'doesn't work'.

In short – the prevailing paradigm has driven practitioners and thinkers into ever more convoluted approaches to attempt to make it work. There is no doubt that the Kirkpatrick idea is simple – in fact its longevity and 'stickiness' has been in this very perceived simplicity. In order to move or remove it, I would contend that a similarly simple idea needs to replace it at its core in order to change the paradigm. However, attempting to apply such certainty to what is, a fuzzy process (learning & impact), and which should be recognised as such, is a challenge for all evaluators who subscribe to such a simplistic approach.

In order to attempt to achieve this, a wide range of subjects were skimmed for suitability to rethink the paradigm and the most pertinent were selected for a deeper examination.

At the beginning of this literature review, I posed a number of questions that the current Kirkpatrick (1954) framework failed to answer. They are

included again with some answers drawn from the conclusions from this assessment of the literature:

Question	Suggested Answer
Why is the framework problematic to implement?	Because it simply does not work in the way it is desired to work in the modern world without a tool-kit to operate it
Why are the results not trusted or valued?	Because the L&D function recognises the nature of subjectivity without the analytical skills to build or understand an answer other than to turn to other metrics
Why is this evaluation process not seen as a vital part of the L&D function given the need to prove results and value	A range of answers from: interest, difficulty, lack of credibility and a failure from the evaluation community to work more pragmatically with L&D practitioners
Why is there so much academic discontent about the framework, yet it still remains as the 'go to' means of conducting evaluation?	Paradigms are 'tough nuts' to crack and the current vested interests are arguably stronger than the desire for change
Why is it still relevant and legitimate in a world so different from the time of its creation?	Because, nothing is better in the eyes of the practitioner field (even though it is recognised not to work)
Does this longevity represent a sign of usefulness or apathy?	At this stage, one assumes the latter.

In no sense can the current paradigm be seen to meet the needs of the world of L&D today; however, it would be remiss to end this review on such a defeatist note – perhaps necessity is the mother of reinvention after all.

2.9 Section Four – New Challenges from External Impacts

It is important for a literature review not just to synthesise the prevailing schools of thought, but also to identify the context and fresh ideas that define the future progression for the practice of evaluation. Section Four aims to situate those ideas into the areas of interest that had emerged from both the literature at this point. However, the primary focus for the ideas in this section are the core of the theoretical base used to construct

the new evaluation framework being developed or that had significance for me as a practitioner.

It is important to contextualise the way in which the L&D function currently operates. In the CIPD Annual L&D survey (2015) it states that a number of trends, innovations and changes are shaping the landscape for L&D functions, for learning within the workforce as well as in the composition of work. A section of these most relevant to the field of evaluation include:

External Driver	L&D Solution or Consideration
Continued need for reduction of transaction cost	The continued rise of blended and e-Learning Improved budget management Greater justification of improvement and the need for value
A realisation of the limitations of 'the classroom'	70:20:10 and blended learning Learner-driven development The rise of Virtual Reality technology
A new generation of learners more attuned to technology than ever before	Innovation in technology solutions New methods of performance assessment
Innovation in work, working life and attitudes to life	Rise in engagement and happiness as well as health and well-being as a theme for both learning needs and evaluation
Continued innovation in the wider learning and development field	Rise of Neuroscience, Gamification, MOOC's, Flip, Hologram and Virtual reality
Increased legal, mandatory and compliance pressures	More compliance and governance metrics required
Need to justify and build the role of HR and L&D to remain central to the changing agenda	Need to measure impact and establish the value of human capital
The lack of 'time' to carry out a complicated process or cope with metrics	More informed and confident L&D functions Better proof of value Faster, simpler processes

In order to service these challenges and respond to this climate, it is my contention that evaluation solutions will need to be:

- Simple, without being simplistic

- Realistic in terms of the slow adaption inevitable within the function because of time and focus issues
- Robust enough to build confidence within L&D and the wider community
- Built on strong concepts, processes and algorithms or even repurposed older ideas
- Relatively simple to convert to an IT platform

It is useful to compare the eternal drivers with the complexity and sterility of the debate within evaluation. Will we, as evaluators, need to create the processes, metrics and tool-kits to help make good decisions and prove value in the new world envisioned by the CIPD (2013), in order to see evaluation as being as important as the 'next new idea'?. It is unlikely that the Kirkpatrick paradigm can answer the questions posed by the challenges above, as it has so far failed to answer the questions posed in the previous section, within the existing world of L&D.

2.9.1 Approach to this section

It is important to recognise that practice spills untidily across fields of theory, schools of thought, and ignores neat search criteria to 'trawl' for ideas that can be re-purposed for the creation of a new evaluation solution. Other researchers may well find different areas of interest that would generate other frameworks and this is recognised as both a strength and a limitation of the end evaluation framework.

In order to build the framework, criteria were set and then reading was shaped by an adapted range of questions that had now changed from the original questions posed at the commencement of the research, including:

- What is the least time-consuming method of evaluation?
- What existing processes within an organisation can be utilised to speed up evaluation and reduce the perception of difficulty?

- What legitimacy can be attached to any new process, to ensure credibility of outcome and confidence in reporting?
- What thinking could be reinvestigated given the transformation that has taken place in technology?
- What new ideas about metrics, process and outcomes are at the edges of the field and appear radically different?

Following the period of trawling across wide ranges of literature, my final choices of focus in order to create the framework were as follows:

- Building on the work of Wang & Spitzer (2005) to investigate Formative methods, to create a concept based on forecasting
- Given the nature of forecasting, the examination of the actuarial, marketing and investment (and venture capital) equity worlds produced useable ideas about the nature of measurement and targeting
- Given the advances in computing power, linking data, Decision Theory to Utility Theory using the work of Cascio (2005) could add value to the nature of targeting
- Building on the ideas in Learning Transfer to create criteria for successful process implementation
- Building on thinkers in Process Improvement and Supply Chain to investigate whether Fuzzy Theory and Grey data could help create innovation in the measurement of intangibles
- Building on the ideas of Anscombe (1957), to consider Intention as a pivotal part of the human dynamic in the process

Whilst each of these subjects could be a doctorate in itself, it is the purpose of this document to give an overview of the key texts that shaped thinking for this framework rather than a complete examination of each of the fields. In this way, future researchers can examine different areas and draw different conclusions.

2.9.2 *Focus on ideas to build into a new paradigm — Forecasting*

"The alternative to thinking ahead would be to think backwards - and that's just remembering"....Sheldon..!

The prevailing paradigm in evaluation suggests that the most reliable method of proving value (often referred to as ROI), is to use summative evaluation. The approach operated by Phillips (1997) is often considered to be the 'gold standard' approach because it is perceived as the most robust because it also used across the fields of accounting and social research as well as training evaluation. However, the central idea for this new evaluation framework was to investigate the concepts of formative evaluation to investigate whether a suitable level of robustness and confidence could be built from forecasting.

Whilst Seigel (2013) contends that there is a distinction between the processes of forecasting and prediction. For this purposes of this document, I have decided to treat the terms interchangeably, to make sense of the many different definitions of each.

Rational Expectations Theory, built on the work of Muth (1961), states that the real world needs forecasting as it is a useful method of making rapid choices about the real world. Muth contends that a rational outcome can be created when certain variables are known and able to be extrapolated. Thus, in supply and demand theory, price can be shown or predicted when the variable of supply and demand can be observed or measured. Whilst opposition to this idea exists e.g. Lucas (1977) forecasting theory has been the bedrock of classical economics since the 1930's. Applying the concept outside of a financial or economic model may be presumptuous, however, it could be contended that the test of a useful framework is its adaptability across fields of thought.

The idea of formative or predictive or forecasted evaluation (summarised as forecasting) is not new. Indeed it is built on sound principles and ideas. The work of an actuary is to establish the value of pensions and pension funds into the future – usually for the lifetime of a person or a cluster of people. Their starting point is to use the idea proposed by Jacob Bernoulli and published in 1713 (later described as the Law of large

numbers), which proposed that, while it might be difficult to predict with certainty a single event, such as the death of a particular person, it was possible to predict with great accuracy the average outcome of many deaths. This principle has built legitimacy over time, because of the accuracy of forecasting and the evidence of accurate outcomes reflecting the forecasts. The members of the profession, whose credibility depends on their ability to calculate pension provision for decades into the future, use a range of forecasting technologies; however, the Bernoulli principle is still used as a core process.

The idea of the standard Central Limit Theory, as reported by Hansen (1982), also bears out this approach, as does a more recent concept of reporting large data averages, reported in the *Wisdom of Crowds* by Surowiecki (2004), where levels of accuracy close to 100 per cent are commonplace from linking large volumes of data and forecasting.

Within an organisation, a vast amount of data is produced, consumed and ignored. The advent of computing power has led to a rise of the concept of 'big data' and many organisations now find the need and ability of make sense of data a competing driver rather as well as a unifying process. This process is often described as analytics. Maisel and Cokins (2014) contend that the purpose of analytics is to simplify data to amplify its value. *"The power of analytics is to turn huge volumes of data into a much smaller amount of information and insight"*. They also separate Reporting and Business Intelligence as a means of packaging the data so that they can be more easily understood by third parties.

Because of the large amounts of data, they contend that most traditional challenges about the 'future' can use reliable data sets to make predictions about future outcomes; for example:

- Which employees most likely to leave
- Which will be the most profitable customer segment
- Increased product shelf opportunity

- Customer Lifetime Value
- Which employee initiative will deliver the greatest return

In fact they conclude that forecasting and prediction have been in the operational managers' tool-kit since the rise of scientific management and are the bedrock of the operation. However, according to DiPiazza and Eccles (1992) "*certain measures can be predictive in nature only when the relationship among value drivers is well understood*". However they recognise that four main methods of prediction have been legitimised – they include:

Method	Overview
<i>Monte Carlo simulation</i>	This process helps in working out ROI when the costs or benefits are unknown, by using the value of any known loss from a worst-case position. The amount and likelihood is multiplied against the opportunity loss from the upside.
<i>Resource Capacity model</i>	Understanding how resources are consumed using (for example) Activity Based Costing.
<i>Delphi method</i>	The use of external experts to forecast an outcome with the benefit of expert knowledge
<i>Scenario analysis</i>	Understanding the risks and opportunities from a range of scenarios

Bayesian theory is also frequently used in forecasting, as the theory changes the core question from: 'What can I conclude from this observation (or what is the probability x is true, given my observation)', to: 'What is the probability of this observation if x were true...'

What matters here is that forecasting is in general use, accurate and has a respected set of tool-kits as part of its theory. Additionally, forecasting is also proposed within the evaluation literature. Whilst Spitzer (2005) worried that whilst forecasting as a concept is useful in evaluation as "*an inexact answer is almost always good enough*", what really matters is the underlying logic, and whether the accuracy of the forecast is appropriate. He cautions the evaluator to understand the limitations of whatever measurement methods are being used and advocates a more 'systems thinking' approach showing how things operate within a context or

environment, utilising: Prediction - what we want beforehand; Baseline - where we start; In-process - what is happening, whilst change is happening, and Retrospectives - after the event. In effect this is a holistic approach and uses a meta approach. However, he did conclude that *"while looking back is helpful, Looking forward is essential"*

Seigel (2013), whilst recognising the limitations of traditional forecasting, built a self-replicating learning loop utilising machine learning to help link computing power to test forecasting assumptions, both to help with prediction and forecasting, as well as to drive enhanced machine learning as well. He contends that *"A hazy view of what's to come outperforms complete darkness by a landslide"*, and that prediction is just 'Induction' or reasoning from detailed facts to general principles which is a basic tenet of the 'scientific method'. Seigel (2013), Armstrong (2001), and Hubbard (2014) all stress that a key test is that all Prediction models have to work on 'back testing' (predicting the past...) taking data in the past to compare results against what actually happened.

Seigel (2013) also highlighted some of the ethical risks in prediction including: Treating people as 'units' of measurement – anyone utilising a measurement approach needs to remember that this tendency can de-humanise a set of data and render any resulting actions less relevant without the appropriate context. In order to counter this, he created an ethical framework to deal with many of his own ethical concerns.

A number of frameworks have been created to challenge forecasting errors in the social sciences. According to West (2006), the following are most useful when dealing with a small number of variables over a short period, whether either simulation or analytical methods are used. These include: Mean squared Prediction Error, (Morgan 1939) and Non Normal prediction errors, Meese and Rogoff (1988) and Misrach (1995), as each concentrates on the link between controllable variables and the use of 'cleaned', large averages to help with forecasting. Armstrong (2001) also recommended the use of 'rival' forecasts and testing forecasts against a

summative process to ensure that the underlying assumptions, logic and technology all deliver a robust outcome.

Walonick (1993) also codifies the different process approaches to forecasting and helps the non-specialist choose and adapt the most pragmatic approach, or to combine them to create a multi-structure approach. In my forecasting framework, the following approaches would be considered:

Approach	Argument
<i>Genius forecasting</i>	Based on intuition and gut feel – <i>rejected due to lack of credibility</i>
<i>Trend forecasting</i>	Useful for short term forecasting but context and 'developmental inertia' can quickly affect the underlying evidence of causality Tools: weighted smoothing methods; turning point analysis; simple linear regression and curve fitting The rise of computing power has helped challenge the idea of Makridakis (1986) that judgmental forecasting is superior to mathematical models <i>Rejected as evaluation forecasting is not trend based.</i>
<i>Consensus Methods</i>	This achieves greater legitimacy, when combined with the Delphic process. The Consensus method is used by Basareb (2007), in his predictive evaluation framework Tools: Appreciative Inquiry <i>Considered as part of the forecasting mix</i>
<i>Simulation methods</i>	This process creates analogies with model outcomes using a computerised approach. Game analogies are used when the interactions of actors are symbolic of social interactions. Walonick (1993) cautions the forecaster to ensure that these mathematical models are accurate at the outset, to avoid the distortion of the simulation Tools: s-Curve; multivariate statistical techniques including multiple regression; Gaming analogies; R-Squared <i>Considered as part of the forecasting mix</i>
<i>Cross impact matrix method</i>	The examination of causality to isolate non-affective variables Tools: Probability, Martingales

	<i>Rejected as being overly complex</i>
<i>Scenario</i>	A set of chosen outcomes based on best, worst and most likely outcomes Tools: Scenario Planning <i>Considered as it is a tool in common use in organisations</i>
<i>Decision Trees</i>	An examination of the structural relationships between problems. Decision theory is based on the idea that an expected value can be calculated as an average. The logic can be refined using Bayes' theorem, and the output can be combined with Utility theory to improve its decision-making process. <i>Considered as part of the forecasting mix</i>
<i>Combining Forecasts</i>	An approach to combine outcomes. Tools: Confirmatory Factor Analysis <i>Considered, but with caution to avoid over-complication</i>
<i>Fuzzy Logic</i>	Swanson (1992) mentions this topic as an area for further consideration. It can be used in conjunction with Random Walk Methods (with or without drift), to embed in a more secure economic framework <i>Considered as part of the mix, however this approach needs substantial access to computing power to enable Fuzzy logic to work – then rejected</i>

One of the paradoxes of forecasting identified within the field is the suggestion that the process may actually help determine a future outcome, perhaps by inadvertently triggering a 'confirmation bias' for example: Modis (1992), Pohl (1993) Dublin (1989). If, as they suggest, the future is discovered, rather than invented, then this could be good news for evaluators, assuming the appropriate ethical framework is in place. At this stage, it is clear that the use of the Delphic and Simulation Methods would be a legitimate and useful means to build the framework

2.9.3 Focus on ideas to build into a new paradigm – Intention

In examining the Kirkpatrick framework (and many others), it becomes obvious there is a need to capture an immediate post-course response from delegates to ensure the highest response rate. Kirkpatrick investigates 'reaction' and, in the Learning Transfer literature, that field identify captured goals. However, I submit that it is more useful to look

at 'intention' as a concept. In other words, what actions the learner actually intends to carry out. This, surely, is the true test of whether the trainer has delivered the actual course aims and actions. In establishing this as a core idea, I leant heavily on the work of Anscombe (1957) to situate the concept.

G.E. Anscombe according to Stoutland (in Ford et al. 2011), was initially influenced by Aristotle and then, more importantly, Wittgenstein, as a teacher of ideas and philosophies and wrote a number of key texts about their ideas. However, she differed with Wittgenstein in a number of ways, rejecting his approach to behaviourism, as well as his appeal for ordinary language. In otherwise following Wittgenstein, her ideas are situated in the 'theory of action' where she represents an alternative stream of thinking.

She contends that any action or behaviour must be as a result of an intention – whether conscious or otherwise. In other words, an output is proof of an intention – likening the difference between potential and kinetic energy in physics. However, she also maintains that the outcome could be different from the intention, but that the intention remained 'true'. In order to explain this, she created a taxonomy of intention, showing the difference between intention states, and that the only way truly to understand the intention linked to the action was to ask the actor their central logic, the question 'why'.

This was different from the ideas of the time, according to Hornsby (in Ford et al. (2011), that were linked more closely to 'logical positivism' where the action was always 'nomological'; where a covering law is used to explain both antecedent and effect. This theory was championed and become a more accepted idea and built upon, possibly most notably by Davidson (1980), with whom Anscombe had many robust academic disagreements.

She differentiates between an intention to 'act' and intention 'with which to act' and how these are similar and different. (articulated further as

'acting with an intention', 'acting intentionally', and 'intending to act'). This was further slightly challenged by Moran and Stone (in Ford et al. (2011)), as a more useful definition of "someone's intending, planning or wanting *to do something* and not, say, someone's wanting or desiring that something or other happen to be the case" (their italics).

In order to understand the intention, the actor must also have 'practical knowledge' as, otherwise, as pointed out by Haddock (in Ford et al. (2011)), "*expressions of intention are distinguished by the possibility of 'mistakes in performance' and not otherwise by their 'direction of fit'*". Haddock (in Ford et al. (2011)) also challenges the notion by claiming that "the actor rarely has knowledge of their intentional actions"

The issues of 'motive to act' and 'intention to act' become a philosophical debate that is at the root of how an evaluator can think about learner aims as, in order to have conscious intentions, certain conditions need to be achieved, to boost the level of intention, including:

- A learner would need to have a degree of (self) recognised competence, built on 'practical knowledge'
- They would need an awareness of themselves
- They would need a level of confidence – in themselves and their competence and
- They would need a level of motivation

Whilst the text and style of writing is problematic in some respects, the clarity of idea and argument is revelatory in its simplicity. Whilst Basareb (2007) uses the word 'intention' as a part of his evaluation framework, it quickly becomes clear that his definition is not part of the Anscombe school of thought and, arguably, has a different level of robustness as it relies on a dictionary definition of the word to denote its meaning. The new framework I will test in my research will incorporate a questionnaire designed for use specifically around some of the concepts in the text from Anscombe.

2.9.4 New Ideas for the Evaluator

What was clear for me from the research, was that the richness of thinking and complexity (and level) of debate, within the academic and practitioner worlds, outside the narrow evaluation literature, were more interesting, enabling and fruitful in terms of approach and inspiration. In my reading, a number of other themes and ideas appeared that could be useful to the evaluator, as the advances in computing power and variation of application have opened up a number of ideas for the evaluator/researcher.

Therefore, in order to inform practitioners, a number of themes and concepts will be highlighted here. There are some ideas that are older and more useful now because of the use of technology, or simply interesting and fresh if used as a new application. These are simply areas of interest sparked in my reading and offered here simply as a means of stimulating fresh thought in the practitioner base; perhaps scattered like 'confetti' into the breeze of curiosity! I will also continue to delve into these ideas post-research, to see what outcomes they may deliver.

Whilst the measure of Net Promoter Score has gained leverage, marketing still has left relatively undeveloped the concepts and metrics of 'Perceived Value' (Monroe 1985). This value definition could be used in training evaluation, by combining the concept with some of the literature around Post-purchase Dissonance, built on the ideas of Festinger (1957) and using the useful diagnostic instrument by Hasan et al. (2104). The use of new ideas from complementary areas has been a rich source of enlightenment in the past and I see every reason to consider other areas from Marketing. In fact, the Perceived Value concept will form an area of interest as part of this research.

Given the nature of the measurement of intangible and subjective information from which an evaluator must draw some legitimate and robust outcome, there may be some future value in dealing with the

concepts of 'Grey Relational Analysis', Chan (2007). This Multi-Attribute Decision making model allows for inference and judgement within a situation or situation that is 'grey', in other words, where many possibilities exist.

One of the methods that can help in this area is Eigenstructure Assessment, Chen (1970). Consisting of eigenvectors and eigenvalues, this promising area of research allows feedback in control loops to affect the system in which they operate. Classically used in hard engineering systems according to Patten (2002. 2014), there is no reason that this application could be used theoretically in service environments in the future, especially with the advances in Big Data and computing power.

The two concepts rely for their academic provenance on the work of Fuzzy systems and Fuzzy analysis. This method allows values to be attributed to more than once constant or number within a set. This concept will be extremely powerful in the area of discourse analysis with the rise of enhanced computing power, by being able to allocate more mathematical rigour to soft variables. In fact, Smithson and Verkuilen (2006) have already completed a number of applications for use in the Social Sciences, and this could be a rich seam of measurement activity as the concept lends itself so neatly to the complexities of human behaviour.

Heuristic Evaluation, Nielsen (1990. 1992. 1994), sits within the area of 'usability engineering' as part of the design process. It consists of methods and tool-kits that, whilst primarily aimed at the front end of a design process, integrate evaluation as part of the iterative process right from the beginning. Independent research shows high levels of efficiency and credibility (Jeffries et al. (1991)) as it involves many stakeholders and forces the issue of design, piloting, testing, and post-implementation review to become 'real time' as well as iterative. Evaluation then becomes tests of understanding and meaning, as well as of perception, and these can then be linked into downstream KPIs, if a more tangible measure is required. The process also requires a number of criteria to be satisfied in order to have greater legitimacy and a likelihood of better

returns. This 'list of usability principles' Neilson (1995) has been constructed by practitioners using some interesting ideas that can directly apply to a learning environment including:

- A reduction in 'scrap' information
- Avoidance of 'interest traps' and other diffusion points
- Language geared to the needs of the learner rather than those of the trainer
- On-going feedback contained within the whole process, to ensure the required output is achieved and not left to chance

This method is well known and respected by IT practitioners and could be utilised by software developers in building a method of evaluation suitable for VR and other blended-learning environments, having much in common with some of the most respected 'Process Focus School' thinkers from earlier in this chapter. Some of these principles may inform the new framework within this research, particularly 'scrap learning'.

The use of technology in the evaluation process is largely driven by the use of the LMS – a process geared towards the reduction in the cost of the L&D function rather than the proof of value (other than its own efficiency). However, other technologies are appearing. Voting and polling buttons can be found in learning environments and evaluators have used discourse analysis on learner forums within Facebook and LinkedIn to generate evaluation outputs. One of the more interesting approaches is the use of an iPhone game to test application of learning which brings together the processes of technological evaluation and gamification. Furio et al (2013) used the iPhone application to create a mixed reality environment within the virtual continuum scale described by Milgram and Kishino (1994).

The game was used both to enhance learning and to reinforce previous learning (which is the principal aim of the learning transfer practitioners), in the area of learning about Spanish Law. Whilst it is recognised that the participants were students aged between the ages of 20-22, perhaps

more attuned to the technology and maybe more likely to be favourable to the use of this type of process, the respondents greatly preferred the game approach to standard learning processes and evaluation irrespective of technology platform. Whilst the authors recognise there is much work to be achieved, they were able to track and measure enhanced outcomes, using an adapted Kraiger approach (1993) on the psychomotor, cognitive, and affective outcomes. Interestingly, the Kraiger framework was also used by Landers and Callan (2012) in the evaluation of learning in Virtual Worlds programmes although this was a small study.

Finally, I suspect that the world of learning and testing will be utterly transformed by the use of genuine Virtual Reality (rather than 360 reality or Augmented Reality). Whilst it is conjectured that Virtual Reality technology applications will not enter the area of L&D for some years in the HR press, it is already being used in the field of medical and surgical training, as well as psychological conditioning. This could be both the greatest opportunity and threat for the L&D function over the next decade. A failure to grasp this area may render the function superfluous.

2.9.5 *Summary and Conclusion*

The future for evaluation is dynamic and hopeful, primarily because of the nature of external transformation that can be brought to bear in the field of knowledge. Whether L&D remains at the heart of the process is a moot point, as organisations begin to flex and remould to take account of the challenges of the nature of work and of the workforce. One imagines that the barriers to technological solutions will continue to reduce, given the nature of comfort with the modern technologies, as the next generation progresses into the workplace.

Within that context, how can Kirkpatrick and his framework survive unless a technological solution makes it work, or we can render the whole idea superfluous, or perhaps if the L&D practitioner embraces a new approach. However, adoption could still be slow because of the nature of

vested interests, perhaps we will have to wait until this current generation is replaced by VR operators or robotic workers!

One final task remains to be completed from the original aims, in that this literature attempted to meet the good practice guidelines, laid down by Gray (2009), and my assessment of progress is as follows:

Coverage	Justification of inclusion in each area is covered
<i>Synthesis</i>	Gaps in current knowledge are covered Evaluation literature is linked to other fields of research and wider issues Background to the issues has been critically evaluated The core themes are linked to the major ideas and issues in the field An understanding of core and peripheral ideas is demonstrated Significance is shown as well as synopsis – tensions and inconsistencies are identified
<i>Methodology</i>	Critical evaluation of past research, with some discussion around previous methods, are deployed in the research
<i>Significance</i>	Research and practice is shown as well as how knowledge has moved forward
<i>Rhetoric</i>	Arguments are created and supported by the structure of the document

At the conclusion of this review, a number of questions, themes, ideas for research and, through the process of reflection, degrees of learning, have been achieved. The ideas for research from this review will be explored in the following Chapter.

Chapter 3: Methodology, Design and Activity

3.1 Introduction & Epistemology

This chapter covers my stance as a learner and researcher and how that stance affected my overall approach to the research, as well as to the decisions and choices I made through the research process. In addition, the chapter will cover the process of choosing research methods with discussion of the implications for the outcomes that were generated. Finally, there will be a discussion of my ethical position that has guided my stance throughout the whole period of research.

A (selected) number of themes had become apparent from the literature review including:

- The degree to which the Kirkpatrick framework still sits at the heart of what organisations describe as evaluation
- A level of personal disappointment of the synthesis of learning between practice and academia
- The (limited) rise of decision theory in the creation of more recent evaluation frameworks
- The realisation of personal limitations in creating paradigms
- The degree of innovation in the fields of thought outside those of the evaluation literature
- Understanding the influences within the evaluation literature on my approach, and the need to reconsider these, in light of wider reading, to create a new framework in the short term for my practice
- Areas of interest, serving as sources of inspiration, both to practitioners and for future research.

These emerging themes shaped some of my thinking and generated questions to surface, during the research process. They will act as a basis and a comparison point, in order to shape questions and to test current attitudes and practice as well as to test whether the emergent meaning, found within the literature over the last 30 years or so, is still current and

evident in practice. They can also help create a point of focus across the life and complexity of the research process to show how the research adapted itself to answer the questions as well as how the questions were re-shaped at each stage to illustrate the required dynamic nature of the research itself.

3.1.1 *Epistemology*

In deciding to consider the scope, limits and significance of learning and knowledge, Creswell (2007) advises that one should consider one's own epistemological stance in terms of what knowledge is and what is possible to know, before determining the technical aspects of any research structure.

One of the challenges in thinking about knowledge is first to understand it. Russell (1912), highlights the processes of 'knowledge by description' as opposed to 'knowledge by acquaintance'; in other words 'knowing how' as opposed to 'knowing that'. This is also reinforced by Anscombe (1959), building on the work of Wittgenstein and described as either 'theoretical' or 'practical' knowledge, and a focus of particular interest to evaluators in demonstrating what is actually learnt from training. This initial idea of knowledge needs to become even more sharply focused when thinking about the role of the practitioner conducting research for an academic audience, or at an academic standard.

According to Fillery-Travis (2012), the practitioner is required to deal with a wide variety of knowledge, in order to integrate and test that knowledge to apply it within practice. Perhaps that leads to a true test of knowledge, in it becoming an application with the outputs generated from that application, rather than for the pursuit of reason? This belief is often a driver of the pragmatic approach, where knowledge is considered to be useful when it works. However, this can obstruct a deeper level of reflection, and be a challenge to a practitioner, when engaging with the methods and processes more common in the world of academic research.

Another issue for me as a researcher is that I work in practice utilising a number of social research methods and approaches. This 'pseudoscientific' world view in approach obscures the reality of truly knowing, seeing that knowing 'enough' is sufficient for our pragmatic aims. Additionally, I discuss, as an evaluator, the 'value of learning' and degrees of knowledge, and the ease and casualness of language can create an over-familiarity with concepts, where a lack of self reflection can leave one open to missing opportunities for more advanced learning and, arguably, even more effective application.

Fillery-Travis (2012) also discussed the 'chain of wanting', where perhaps 'the search for knowing in practice' develops the motivation to engage with the academic world, and a range of positions are developed to help with the creation of reflective practice. Understanding what and when knowledge is 'enough' has driven my 'chain of wanting', in order to engage with a more deeply reflective approach to learning.

Lincoln (2001) discusses learning as the "*frame for judging what may be known about the world, and the relationship of the knower to that which might be known*" and, as someone that believes that the world of academia and practice have much to offer each other when each can offer perspectives to the other, it is useful to consider personal epistemology in order to challenge ones own habitual world view.

Also, considering the impacts and ethics, as both a practitioner and a researcher, I need to be sure of my stance with regard to my self and my practice in both areas. In order to help with this, it is worth examining some core concepts, and how they sit with the perception of my approach.

In considering the process adapted by Crotty (1998), to determine relationships between the researcher and research phases, four headings allow for reflection and choices to be made: They are:





In reflecting on the Epistemology stage of the framework, and the three sub-choices of Constructivism, Objectivism and Subjectivism within it, Grey (2009), challenges the researcher to consider how to attribute meaning to the world around the learner, by considering approaches to reality. I would consider my epistemological position to that of a constructivist, as I believe that truth and meaning is created with interdependent interactions in, and with, the world. In doing so, I therefore reject the position that there is an Objectivist reality somewhere 'out there' to be uncovered in a world separated from myself – surely that must be created by someone for something? I also reject the idea of Subjectivism, where truth is discovered through the interplay within a collective, unconscious being "represented by dreams and religious beliefs" (Gray, 2009), as this defies for me any sense of logic as a pragmatic means of finding truth or knowledge-building.

Having identified an epistemological position as a constructivist, this will now assist in the cascade of decisions, as shown in the Crotty (1998) process, and help lead to the creation of knowledge, as I now know I need to construct truth and meaning from interactions. Therefore, in considering a 'Theoretical Perspective', to build on my epistemology, I realise that Interpretivism offers the broadest and most useful perspective in helping structure an approach to deal with the real world issues in hand.

In coming to this conclusion, I considered, and rejected, positivism, as it lends itself initially to the concepts of deductive and inductive research, and Crotty (1998) suggests that it implies that research outcomes '*will tend to be presented as objective facts and established truths*'. However, because of the nature of the subject matter of the research, and the complexity of the range of theory and practice, this would create an issue in being able to meet the positive principles, outlined by Bryman (2008), of '*pure empiricism and factual logic building*'. Those methods are too

removed from the complexities and interpretation required from research in the wider field of social sciences.

Interpretivism also contains number of sub-themes that can be selected to fit with an epistemological position and around which to situate research methods. These include: Symbolic Interactionism – this has some appeal as, according to Gray (2009), Symbolic Interactionism effectively links social meaning with behaviour, stating that *"meanings arise from the process of social interaction"*. Given the nature of evaluation, as having a range of actors, then one approach is to consider the idea of learner-centric evaluation strategies, this approach could have value, as the approach is predicated on the idea that meanings are fixed on the basis of actions. According to Blumer (1969), this involves *"catching the process of interpretation through which actors construct their actions"*. However, this approach had to be rejected, because it does not allow a researcher to create an organisation-wide process, as it is too closely linked with the meanings and truths of too few people within the evaluation process.

Another approach within Interpretivism is Realism, which is the idea that a reality independent of an observer exists to be discovered, was also examined and discarded on the basis that it relies too clearly on the idea that there is an external reality 'out there' that can be discovered and that new knowledge is an extension of the existing reality. This appears to have more in common with an Objectivist approach and would imply huge leaps of faith in imagining that there can be an objective view of intangibles such as 'culture' etc. At the other end of the scale, sits Naturalistic inquiry with a view of multiple, constructed realities that need to be studied holistically. Guba and Lincoln (1994) suggest that prediction and control of outcomes *"is a largely futile expectation"* and, whilst being curious at the apparent internal chaos of the approach, I discarded this on the basis of a lack of rigour and credibility to a sceptical practitioner-base, of any result that may appear from this approach.

Further approaches also include: Critical Inquiry, (the idea of defining and changing 'false consciousness'); Postmodernism and Feminism were also discarded, based on the extent to which the approaches create a rigid framework that imposes a set of 'values' on information or data. Each approach seeks to change the outcome of the subject, through the questions that are being posed, and I decided this would add little value in an area that was already confused with complexity and vested interests. It could be argued that these approaches require the researcher hold a particular world view or position through which the research would be interpreted, and I wanted to avoid this in order for a creative process to be possible later in the research.

Hermeneutics, yet another Interpretivist approach, considers social reality as being "*socially constructed rather than an objective fact*". Whilst hermeneutics initially focused on the interpretation of scripts and literature (often sacred), according to Grondin (1994), it has widened to areas of broader or more general interpretation and the teaching of those interpretations. Whilst, initially attracted by the approach, it was discarded on the basis of the interpretation being predicated on the ability of the researcher to have deep levels of knowledge and self-understanding. In addition, the role of the researcher would assume too great an emphasis in the voice of the research, rather than constructing, or interpreting, meaning from the culture and views of a wide range of informed opinion.

Perhaps unsurprisingly, Phenomenology (the idea of discovering meaning from emergent phenomena), became the most appropriate research approach, as it has to be grounded in both social reality and the experiences of people within that social reality. In other words, people reflect the world in which they live and operate. For example, organisations are deemed to have within them a culture. Becker (1975), illustrates the view of culture as "*people create culture constantly....and adapt their understandings of what is different about it*", and this view can be shaped by actors and creates a rationale for change, based on the

appreciation of skill and influence of the actors within it, therefore, the culture can be identified by the actions of the actor and allows meaning to be interpreted from the actions of the actors. In addition, phenomenology, in surfacing the idea of phenomenon, allows the researcher to identify and include a large amount of data and develop understanding from a wide range of conceptual and observational information, from which can come meaning.

In order to achieve this, according to Gray (2009), current understanding of phenomena have to be 'bracketed' to allow those phenomena 'to speak for themselves'. The role of the researcher is therefore to look at, and find, new meaning or perspectives in the phenomena. This approach allows the research to become more full of 'meaning', which can be used to 'construct' knowledge and truths, sometimes by using an inductive process to make sense of the data. As large samples of data can be collected, this allows the researcher *"to pick up factors not part of the original research focus"*. The nature of the project, the breadth of actors engaged, and the realities of the nature of the theories and subject area, suggest that an Objectivist view of Ontology would have been less than useful. No aspect of the project can be seen to be served well with a paradigm where actors within it cannot shape the world in which they operate.

One of the risks of the interpretive approach is within the role and worldview of the researcher. The degree to which the influences, values and prejudices have been eliminated, as meanings have to be interpreted from the project, especially where vested interests in seeing a useful outcome need to be considered is vital. Also, Bryman (2008) talks about the need for the researcher to remain capable of being surprised by the findings, the need to suspend judgement too early in the process will be a vital part of not leaping to conclusions, or seeing the phenomena as having the possibility of bearing different meanings.

An example of this approach to 'bracketing' will be the reflection needed to find the meaning from each of the phases of any research; especially

as the nature of the research may well find new ideas not known at the beginning of the process. Having the personal discipline to follow the research process, if the research doesn't seem to be producing phenomena with sufficient meaning, can be a challenge for a researcher to consider if, or when, that situation arises.

Developing self awareness will be key to this process and, in order to help me become more accomplished with this idea of 'bracketing' the phenomena, I decided to build on the work of Costa and Kallick (1993) and acquire a 'critical friend', to discuss all results and findings in order to check my assumptions, judgements and potential lapses of objectivity. A PhD herself, she can act as a brake on the commercial realities of the project and mediate between my various enthusiasms, ideas and initiatives where required. Her role will also be to ensure that the level of surprise articulated by Bryman (2008) is not overwhelmed in the academic process or the pragmatic needs of practice.

As I have read more and become more reflective about my own learning and route to knowledge, it has become increasingly obvious that the Patton (2002) description of a pragmatic view of the interpretivist epistemology particularly suited me and fitted well with my stance as a practitioner with real respect for academia. All in all, I had much in common with the realisation by Peters (1992) about the need to find both a subjective and objective stance that allowed one to maintain a healthy regard for the practitioner clients in wider practice, as well as those within the framework of the research. Having become more comfortable with the drive towards a pragmatic approach, the research structure would incorporate a pragmatic response to both problems and opportunities.

3.2 Research Questions

In conducting the literature review, it had become clear to me that some of the research questions with which I had begun the process had changed and developed in the context of wider thinking, exposure to the fields of thinking, as well as some of the potential advances in terms of innovation in day-to-day practice. My initial, grandiose aims of paradigm replacement were now tempered with the reality of the researcher in seeing their place in the scheme of things and that my world view needed to focus on the establishment of good research and outcomes, allowing the view of others to decide if this was worthy of a paradigm shift.

In constructing a series of research questions, I reviewed my original list and realised that I had been thinking in too tactical a manner and that, guided by some of the themes identified earlier in this chapter from the literature, that the new questions should be as follows:

- How can a new evaluation framework, built upon Decision Theory principles, add value and allow L&D to prove value in a cost effective way?
- How can the competing drivers for the need for change and the need for credibility in evaluation be resolved in a function under pressure?

A range of sub-questions were refined as follows:

- What new or existing attitudes exist in the practitioner base that continue to drive or restrict attempts to evaluate?
- Does the outcome required by the stakeholder create some of the confusion at the heart of the process?
- What part does credibility and resource effectiveness have to play in any solution?
- Does the longevity of the Kirkpatrick framework represent excellence, apathy or a lack of choice?
- Should the process of 'what is possible in technology' create the process of evaluation that organisations buy into?

Other, more tactical questions, are posed as part of individual chapters, and serve to structure the process of thinking as well as to help the progression of the research and subsequent outcomes. In order to answer these questions required a number of research phases to be considered. These will be discussed in greater detail later; however, the outline for consideration would be:

- A broad review of attitudes to and about evaluation from a 'credible' population of learning practitioners
- A deeper review of the current practice from a smaller population of evaluation practitioners
- The development of a framework to use and from which feedback can be gained and data collected
- A review of outcomes and resultant attitudes to evaluation, based on the data collected from the activities in the previous phase.

My initial position with regard to the questions was as follows:

Main Question	Discussion
How can a new evaluation framework built upon Decision Theory principles add value and allow L&D to prove value in a cost effective way?	Whilst recognising that evaluation is made from both processes and measurements, this does not mean it cannot serve a grander, or more strategic, purpose by utilising Decision Theory. The idea that evaluation could become more meaningful, as well as cost effective, would allow a more strategic, value-adding approach to be enabled. The proposal to create a framework is an important part of practice, and to examine ideas gained through theory and knowledge. The framework can also test out some of the more innovative approaches and perspectives gleaned in the literature review to help think differently about the first question
How can the competing drivers for the need for change and the need for credibility in evaluation be resolved in a function under pressure?	The need to build credibility in the mind of the practitioner will be driven by results, both in terms of outputs, resources requirements and costs. Linking to strong theories will build credibility in any area of innovation.

The position to the sub-questions were considered as follows:

Sub Question	Discussion
What attitudes exist in the practitioner base that drive or restrict attempts to evaluate?	The attitudes in the practitioner based affect both the creation of a new paradigm as well as the extent to which tactical evaluation is a problem for current practice. Whilst there are examples of research that focus wholly on this area of research (attitudes) for an entire document, this is viewed and treated as a simple starting point for this research
Does the outcome required by the stakeholder create some of the confusion at the heart of the process?	The literature suggests this to be the case, rather than being a problem wrought by practitioners. Comparing the attitudes of practitioners will determine how much the stakeholder base actually matters to them
What part does credibility and resource effectiveness have to play in any solution?	Whilst this area seems simplistic, it may well be that the practitioner base is more troubled by their own workload than the whole strategic imperative, as the literature suggests the move to a service delivery role. Whether the new approach lightens or further complicates resource considerations would affect attitudes and subsequent implementation.
Does the longevity of the Kirkpatrick framework represent excellence, apathy or a lack of choice?	Does the Kirkpatrick method really represent the pinnacle of evaluation in practice? Attitudes and access to new information will be a key part of this sub-question. It would be interesting to see whether a function so well known for innovation, 'the next new idea' and the pursuit of the boundaries of learning (CIPD 2014) would have the same approach with evaluation.
Should the process of 'what is possible in technology' create the process of evaluation that organisations buy into?	As many evaluation solutions are actually technology driven – maybe it should be the IT function that determines the evaluation strategy?

3.3 Discussion of Methodology

Before beginning any research, it is seen as vital to select a research method that reflects the epistemological approach that was discussed

earlier. My position for the research was to decide that an inductive approach was most useful. Whilst a deductive approach may arguably lead to a more robust solution through the statistical 'proving' of a hypothesis, the research needed to be more flexible and informed and shaped by the views and opinions of those most likely to use it. Also, whilst much is known to me about the theory of evaluation and the views of the L&D profession, it seemed reasonable to put aside my narrow experience and to investigate what is current practice, as well as a deeper investigation of practice within a more expert (user) group, by using an inductive approach.

However, on reflection, I realised the research will need to consist of a mixed methods approach, in order to reflect the different stages of each approach.

Stage	Actions
Accumulation of facts, data etc.	Assess current knowledge from my own practice Undertake a literature review of theory Run two research processes to test wide opinion from a large sample Run a survey for a 'narrow' group of expert practitioners
Analysis and validation or comparison	Analyse data and match to known data
Formulation of the framework	Assess literature from parallel areas of expertise Make an 'intuitive leap' through the analysis of available data, frameworks, and a process of innovation to synthesise into a framework
Application, Monitor & Reviewing...?	Create a case study group and monitor the application

Describing the research in the manner above may lead one into considering that the overall approach looks suspiciously like Action Research rather than the mixed methods approach outlined earlier. I had also been confused about the point and that was reflected in my

searching for the most appropriate research tactics, until I was sure of my stance.

I considered a number of Gray's (2009) strategies of enquiry and discarded the Action Research and Experimental approaches in order to focus on a Case Study, because of the lack of iterations of the process of research. Also, the idea of a bounded system proposed by Stake (1994), where one can test an idea within a series of limited constraints, but with the flexibility to apply mixed methods, became the most appealing. This was partly because I had access to organisations that were happy to be treated as such (case studies), as well as the structure benefitting a more blended approach. This also met my need for both qualitative and some quantitative rigour in methods, and in order to create practical legitimacy as any results situated in the 'real world' of the practitioner base tend to have greater credibility, rather than using theoretical evidence. However, that being said, according to Pawson and Tilley (2001), the 'gold standard' of research is pluralism, where methods are used according to opportunity and need. Perhaps, in applying my pragmatism this makes more sense to me.

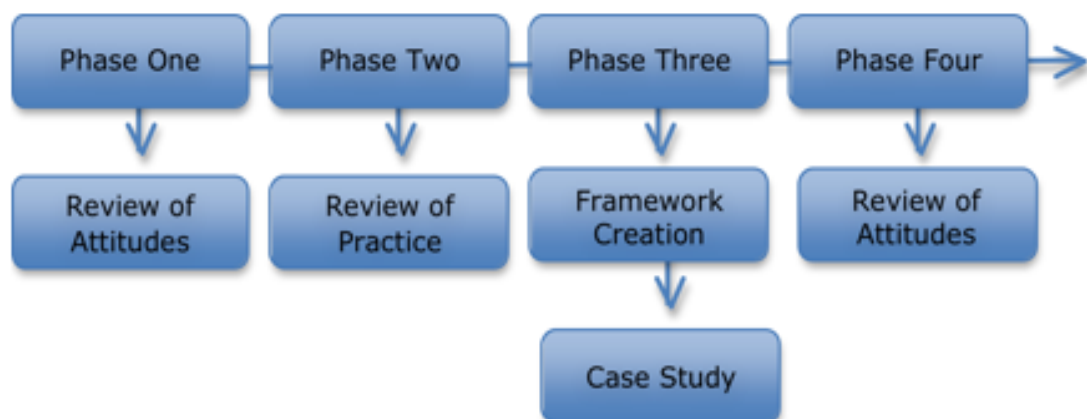
So my decision can be presented using the Saunders et al. (2012) approach, presented in Gray:

Epistemology	Constructivism
Theoretical perspective	Interpretivism
Research approach	Inductive & Case study in appropriate phases
Time-frame	Cross sectional
Data collection methods	Sampling, Observation, Interviews, Questionnaires, Review of organisational data

3.4 Research Design Overview

Having become clearer on the research questions, overall approaches and stance of my research, the next sections outline, and illustrate, the choices and decisions for each stage of the actual research. Initially, the research had the grandiose aim to replace an existing paradigm; however, it became apparent during the research that this idea had to be moderated, to create simply a framework that could become a paradigm if it had sufficient authority, practicality, and credibility. The initial single phase of research therefore became expanded into four more robust phases, each of which had its own challenges and methodological considerations. Each phase will be discussed in turn:

Overview of Research Phases



To give a little more detail:

Phase	Design	Deployment	Analysis
Phase One	<u>Sample</u>	<u>Survey</u>	<u>Content Analysis</u>
<p>2010 – 2012</p> <p>'Initial review of learning practitioner attitudes to evaluation'</p>	<p><i>Purposive Sample Stage 1 – invitation to contribute – verbal invitation – 708 acceptances</i></p> <p><i>Purposive Sample Stage 2 – Group reduced to 436 contributors</i></p>	<p>Questions distributed as follows:</p> <ul style="list-style-type: none"> - 314 Online questionnaire responses - 28 one to one interviews - 51 using telephone response - 43 paper survey completions 	<p>Content analysis using proprietary software – converted to excel for reporting in this document</p>
Phase Two	<u>Sample</u>	<u>Survey</u>	<u>Content Analysis</u>
<p>2014</p> <p>'Review of smaller population of evaluation practitioners to assess current evaluation practice'</p>	<p><i>Purposive Sample Stage 1 – 65 companies</i></p> <p><i>Purpose Sample reduced to 43 organisations</i></p>	<p>Questions distributed as follows:</p> <ul style="list-style-type: none"> - 24 Online questionnaire responses - 6 one to one interviews - 13 using telephone response 	<p>Content analysis using proprietary software – converted to excel for reporting in this document</p>

Phase Three 2014 – 2016 Stage A – Framework Creation Stage B - Case Study	<u>Stage A Framework Creation and Sampling</u> Framework formulation - Stage One Ideation – reading and concept formations - Stage Two - Testing of the concept – workshop of 12 experts Creation of initial forecasting accuracy benchmarks <u>Stage B Purposive Sampling</u> to select Case Study companies – 7 companies with 6 L&D practitioners	<u>Stage B cont.. Process Operation</u> Expectations of forecast accuracy collected from Case Study group Retrospective process trialed on 27 courses where benchmark information existed New framework process run in 7 companies and over 94 courses Third party 'gold standard' process then completed in same companies and same courses to create the benchmark score	<u>Data Comparison</u> Third Party results using 'gold standard' process results turned into the 'benchmark' against which the variance of the forecast is measured Cost profile created Outcome comparison
Phase Four 2016 To review the attitudes of those taking part in the Case Study	<u>Sample</u> <i>Purposive Sample of all 6 people who were the key L&D practitioners involved in the process</i>	<u>Survey</u> Questionnaire sent to all 6 respondents. All 6 also interviewed by a third party	<u>Content Analysis</u> Content analysis using proprietary software – converted to excel for reporting in this document

The following sections explore each of the phases in greater detail.

3.4.1 Research Phase One – Poll of Attitudes

Phase	Design	Deployment	Analysis
Phase One	<u>Sample</u>	<u>Survey</u>	<u>Content Analysis</u>
2010 – 2012 'Initial review of learning practitioner attitudes to evaluation'	<i>Purposive Sample Stage 1 – invitation to contribute – verbal invitation – 708 acceptances</i> <i>Purposive Sample Stage 2 – Group reduced to 436 contributors</i>	Questions distributed as follows: - 314 Online questionnaire responses - 28 one to one interviews - 51 using telephone response - 43 paper survey completions	Content analysis using proprietary software – converted to excel for reporting in this document

Introduction and Selection

In my original thinking for this research, this Phase was intended to be the major focus for the whole project and, as such, I wanted to build a degree of 'credibility' into the findings by asking a small number of questions to a relatively large sample of people to draw out key attitudes to evaluation.

I considered a range of sampling strategies in order to begin to create the concept of credibility including: Random sampling, including stratified, cluster and stage sampling. However, it quickly became apparent that these methods were both time consuming and restricting the actual numbers of people I could talk to. 'Cold' approaches from external parties looking for research data can swamp L&D functions and

these appeared to be inappropriate for my aims, as I had plenty of opportunity to interact with the target sample on a regular basis.

Also considered were Non-random sampling, including Quota, Convenience and Snowball sampling. None of these methods lent itself well because of the access I had in the many large networks available to me in the practitioner base more usefully than the process of 'Purposive sampling', which was the method finally settled upon. Whilst the risk of this approach is that a researcher can miss a vital characteristic or be subconsciously biased in the process, I considered that this a manageable risk against the benefits of being able to research with a credible cross sample of the L&D population.

My initial strategy was to talk to every single respondent, or to survey those who had already verbally agreed to take part in the survey so that the actual response rate would be 100 per cent. Therefore the initial stage was to collect commitments from people to take part in the research. This is unusual in research, but was possible for me because I have a large number of opportunities to interact with the research base through my specific consulting activities in this area. I work within a specialist evaluation company (one of the very few in the UK), and have access to clients and sensitive data in this area, where a more generalist consulting firm may be regarded with greater suspicion.

So we are regularly able to meet and access people and elicit opinions:

- We meet them at conferences and exhibitions
- We present ideas and solutions to the research base at both online and offline events
- We interact with them through marketing events and during pre-sales meetings
- We belong to networks and groups where we have the ability to pose research questions
- We have clients in this field with whom we interact directly as well as being able to access their third party contacts

- We have access to training practitioners and their third party through the work we do with them on behalf of clients

I began the process without a clear idea of how many people to contact so we asked people who were a) interested in the subject and b) prepared to take part in an academic research project.

The initial question asked was “would you be interested in taking part in some research about the future direction of evaluation of training in order to replace the Kirkpatrick paradigm”. Those that answered ‘yes’ were then part of the process.

I asked that question over a period of some nine months and then paused so I could review what had been collected. At this stage, I had a range of commitments from over 700 people. Now I could begin to refine the sample and develop the actual questions for the Phase One research.

As part of reviewing the commitments collected, it became obvious that such a large research base had gathered too wide a range of perspectives and interests – for example the opinion of an external training company is driven from a different perspective from the target sample of in-house L&D people. I also decided that non-L&D people were, in fact, a customer grouping for the L&D team, and whilst important, this review is aimed at, and for, the practitioners in L&D.

I therefore decided to narrow the sample and reduce the total amount by *removing* the people in the research base that fell into the following categories:

- Training Consultants and Trainers
- Any operational manager not in HR or L&D
- Any non-manager in L&D and HR (e.g. HR adviser or L&D administrator)
- Any non UK company (or the non UK element of an international company)

- Any LMS supplier
- Any software supplier (evaluation software or otherwise)

The resultant group consisted of 436 people and consisted purely of Managers, 'Heads of', and Directors of L&D from a range of companies in the UK. The issue I had to resolve was whether this number was, as recommended by Gray (2009), legitimate, robust and sufficiently randomised.

I applied the following justifications to the criteria to test out whether that sample size would be credible.

Criteria	Justification
<i>Robust</i>	Each person was the correct level for the research project. The reduction in the number had ensured that irrelevant job titles were eliminated and that only people with relevancy were included I consulted with my learning mentor for advice and guidance
<i>Legitimate</i>	The actual number of respondents was larger than many of those in other similar research projects; however, I was concerned that a small number of questions would add insufficient enlightenment and create data saturation, therefore, I elected to use more open-ended questions, to give more scope, and a more detailed response.
<i>Sufficiently randomised</i>	The initial process of selecting people had been completely random – based on those simply interested in the subject. The removal of job titles had served only to tighten the relevance of the remaining group. There was an acceptable spread of companies in terms of sectors and size.

3.4.2 Phase One – Research Data Collection

Aware of the need to capture opinion as a key part of this process, and aware of the fact that the majority of interactions would not be a traditionally randomised survey, the bulk of the questions were designed to be open-ended so that themes and opinions could be captured. This fitted with an idea of the descriptive (rather than analytical) survey in Gray (2009). The questions that were asked are shown in Appendix 3.

Taking heed of good practice, the questions were tested on a small group of colleagues, and my learning mentor, to ensure the questions were both understandable and would elicit the data I needed to collect. Following some minor amendments, the questions were then asked in a variety of ways and driven by the requirements of the respondents:

- In a series of one to one interviews
- Through responses to online surveys
- Over the telephone or social media, where required
- By completion of paper documents during events or conferences

I decided that the process would have legitimacy because although the 'channels' of data collection would differ, the questions would remain constant.

One of the biggest challenges in the small number of face to face meetings was the subject of 'questioner bias' According to Gray (2009), this is a challenge and I instituted a review process by using a colleague to 'sense-check' the outcomes I was collecting, to ensure no supplemental questions were being asked or the subjects were becoming too diffused by my own curiosity. Each response was written down verbatim and put into a dedicated database for analysing open-ended questions (online contributions were fed directly into the database). Aware of the risk of short-cutting this process, a manual quality check was run with each thirtieth form being checked by a colleague for accuracy of input.

3.4.3 Phase One – Data analysis methods

Dey (1993) suggests that qualitative analysis is a circular process of: Describing; Classifying, and Collecting and this idea runs through each of the stages and the Phases outlined later.

In designing a set of questions to find out about attitudes, this produced a commentary wider than that initially aimed for in the questionnaire. This wider discussion and commentary was captured and analysed as part of the whole. This also helped shape some of the additional questions that were in the next Phase (Two) of the process.

The method chosen to analyse the data is to use Content Analysis, rather than Grounded Theory, as there was no need to build theory from the 'ground up', as a wealth of theory and information existed already and the focus is to identify the present situation against which meanings can be generated.

However, Flick (2006) suggested that a challenge in Content Analysis is to ensure it remains clear of confusion about the nature of the content, in order to ensure that each step is considered separately as well ensuring all the steps have been considered. In this way, a richer analysis can be generated.

These steps are as follows:

Steps	The challenge in the process
Summarising content analysis	<i>Where content needs to be paraphrased because a person has used the same response, anecdote or idea to illustrate different questions.</i> Not to do this would create an undue emphasis on specific ideas or themes. The challenge here is the problem of 'expert knowledge' and ensuring that content is not polluted with personal interpretation of the summaries.
Explicating content analysis	<i>Where content needs to take account of organisational cultural norms, slang, acronyms or other jargon.</i> The

	challenge here is to ensure understanding of the nature and context of the organisation in hand.
Structuring content analysis	<i>Where the content has been used to create scales (arbitrary and subjective) to assess, for example, strength of feeling for a particular idea of attitude.</i> Whilst subsequent searching for 'strength of feeling' indicators can be carried out, the level of individual subjectivity makes this a worthless exercise. The challenge here is to realise that these scales can never be comparative in any more than a subjective opinion.

The process, in practice, to produce the analysis is as follows:

- Initial physical reading of the survey results
- Coding of the responses – firstly using repetition analysis then themed content analysis
- Input of the codes into the database

The coding process has a number of practical challenges that need to be overcome:

Challenge 1 – To review the content or the respondent?

Respondents often make multiple points and a decision is made to review the data rather than simply the initial response. This means a respondent can often offer a substantial amount of content to a question where they may have strong views and could skew the data. The risk of this happening is preferred to the subjective inclusion of a single point made (e.g. the first point), by the respondent per question.

Challenge 2 – To analyse the words or the meanings?

Respondents often reply in a manner that can be misinterpreted if not coded by a skilled practitioner. For example, if deciding whether the following statement is a 'positive' or 'negative' attitude, is about using skill to apply meaning to the statement: "Our appraisal system is excellent – not that I have ever had one from my current manager". Using calibrated judgement is a strength of the approach and allows for both the collection of the main attitude or response, as well supplemental information.

The coding process of the analysis stage takes an average of eight minutes per response in Phase One, as there are few questions and less ambiguity than usual. The outputs are produced very quickly from the software, which translates them to online interactive dashboards and documents. However, due to the nature of this research process, this is unrealistic and, therefore, the outputs have been translated into spreadsheets to create graphs for third party consumption. Our data storage and confidentiality processes are shown in the Ethics' section 3.9.

3.4.4 Phase One – Conclusion

Phase One ran smoothly and generated large quantities of data and some information that added to my understanding of new and existing meaning from the wide base of people whose attitudes had been collected. Within this data were some surprising findings around the longevity of the Kirkpatrick framework and of some of the pragmatic reasons why it still existed and would remain in place for some time to come without a more meaningful and credible alternative.

Phase Two would focus more on the contribution from a specific 'expert' evaluation base, about current actual practice and the degree to which they were using new approaches as well as to compare findings in the literature with their current practice.

3.4.5 Research Phase Two – Evaluation Practitioner research

Phase	Design	Deployment	Analysis
Phase Two 2014 'Review of smaller population of evaluation practitioners to assess current evaluation practice'	<u>Sample</u> <i>Purposive Sample</i> <i>Stage 1 – 65 companies</i> <i>Purpose Sample reduced to 43 organisations</i>	<u>Survey</u> Questions distributed as follows: - 24 Online questionnaire responses - 6 one to one interviews - 13 using telephone response	<u>Content Analysis</u> Content analysis using proprietary software – converted to excel for reporting in this document

Introduction and Selection

One of the challenges faced by a researcher and their chosen epistemology is the time lag between concept and execution, especially where there is an evolving number of ideas and themes to interpret. The original idea within the research became superseded by the need for further phases of research to generate information, in order to generate further insight. The initial shape of the new framework was developing as findings from the literature review, and wider reading began to stimulate new ideas.

Phase Two therefore became focused on current practice of a smaller purposive sample of evaluation practitioners, who were involved in evaluation within their L&D function, in order to help add value and insight. Whilst another, more detailed review of attitudes may have created more insight into the wider practitioner base, I was aware of the risk of data saturation and wanted a more pragmatic comparison of practice and the literature. In addition, the decision to carry out a review of the accuracy of the new framework suggested the need for a Case

Study and a review of practice would offer insight, and would also allow the synthesis of learning for this process.

Therefore, the focus of this Phase was to examine:

- Attitudes of evaluation practitioners with greater experience and involvement in evaluation than most of those in Phase One
- Current practice around new approaches to evaluation metrics, as well as processes around evaluation and learning transfer
- Attitudes to the current use of what would later be described as Decision Theory
- Collection of data and information to inform a new approach, particularly to collect evidence about specific ideas (for example, formative evaluation, learning transfer etc.)
- Case Study Suitability – some of the evaluation practitioners completing the survey in this Phase were interested in helping test the new framework as I began to socialise the idea of it and, as they heard me speak at conferences and on webinars. Therefore, using some of the data helped to assess who had many of the necessary processes in place and became a way of screening those who wanted to take part.

In order to help identify the overlaps in the process between Phases One and Two, I created the following diagram to help identify the assumptions and differences, at each stage, so that the methods' section could be clarified and to ensure no steps were missed out or trivialised.

Stage	Phase One	Phase Two
<i>Sampling</i>	Purposive and chosen from access to the formal and informal networks in the practitioner base. Sample further refined against criteria	Purposive and selected mostly from the initial research base with the addition of one or two extra interested parties – all were evaluation practitioners
<i>Question Design</i>	A small number of mostly open-ended questions to elicit broad attitudes.	A large number of questions to investigate practice and to create

	Process piloted with a few amendments	insight. Also the phase used to select the case study participants. Process piloted with many amendments
<i>Data Capture</i>	Surveys split between semi formal and semi structured questionnaires	Online surveys with some face to face interviews to facilitate respondent needs regarding time pressure or disability (one person was blind and preferred a conversation)
<i>Data Analysis</i>	Computer-based content analysis used to generate results Initial review of the data generated primary coding. Results exported to excel spreadsheets for the document	Computer-based content analysis used to generate results Initial review of the data generated primary coding. Results exported to excel spreadsheets for the document. Some statistical analysis completed in addition to understand the strength of themes etc.
<i>Data Presentation</i>	Excel graphs from exported data	Excel graphs from exported data

Phase Two of the process initially involved a sample of 65 companies. I had initially identified 'expert' volunteers from the narrowed-down sample in Phase One. As a result I had 55 of those that had indicated interest, to which were added ten existing clients of my consulting practice. The idea within Purposive sampling (with a healthy dose of Pragmatism) was again the key technique in determining both the constituency size and nature of the sample. At this stage I decided to narrow down the sample group for a number of ethical and practical reasons.

Therefore I removed any NHS, Local and Central Government organisations from this research phase, as these organisations have particular issues and challenges (both political and regulatory), in working with external consultants and have complex ethical drivers that

would have precluded them from taking place in the resulting pilot of our new process.

The remaining group had 42 people/organisations within it. The criteria for their inclusion were:

- They were conducting evaluation in some form or other
- They were interested in the subject and wanted to be part of the research
- They were interested in seeing the results of the research
- They had a variety of evaluation issues – from a single programme to a total budget and hoped that seeing the outputs would help them in some way
- They needed a credible approach to evaluation

I was again concerned about the resulting sample size and decided to re-run the process I had used in the previous research stage to resolve whether this number was legitimate, robust and sufficiently randomised.

I applied the following justifications to the criteria to test out whether that sample size would be credible

Criteria	Justification
<i>Robust</i>	Each person was the correct level for the research project. Each person wanted to be part of the research and had relevant practice to share A number of 'experts' were in the group – particularly those who had been trained in evaluation methods from competitors of ours I consulted with my learning mentor for advice and guidance
<i>Legitimate</i>	This was a purposive sample of expert practitioners with a range of experience and expertise in the subject. Also every person in the sample contributed, thus giving a 100 per cent completion rate
<i>Sufficiently randomised</i>	The initial process of selecting people had been completely random in Phase One – this process step

	was also randomised by the inclusion of extra practitioners who joined in late, and further randomised by the level of interest of the individual concerned rather than the specific sector or size of organisation.
--	--

3.4.6 Phase Two – Data Collection

Having considered Phase One to have been successful, the same process was operated in this phase. Ethical statements were agreed and each person was guaranteed anonymity (whilst a number were happy to waive that anonymity, I considered it important for those who wanted it to ensure that the whole group operated at this level of ethical protection). A wider discussion of ethics takes place in section 3.9.

The majority of the group (33) had indicated they would prefer to complete an online questionnaire, therefore this was created in order to be administered through Survey Monkey, for those people that preferred this type of interaction. The remainder (10) received the questions in advance and their answers were captured verbatim and then input into the database, for analysis.

The questions were designed and circulated to a pilot group of colleagues (6) within my own practice for testing.

I received much more challenge for this stage particularly around:

- The number of questions in the document – concerns were expressed about questionnaire fatigue
- The repetitious nature of the questions (particularly in the specific need for detail I was seeking)
- The level of specialist contribution required from participants
- The blend and of types of question (I was asked to group questions together to ensure there was a clearer linkage between themes. I had been keen to mix up the themes to avoid leading the respondent down my 'train of thought', but I was reassured

that having stronger themes enhanced understanding about the different subject areas that were being considered).

All this was taken into account and built into the briefing given to participants. The briefing outline is contained in Appendix 3a. As there was no deadline, the results took around 4 months to collect until the final returns were received.

3.4.7 Phase Two – Data Analysis

Using the same processes as in Phase One seemed appropriate to ensure there was consistency across the Phases of research in the effective discovery of meaning. However, with a wider blend of open and closed questions, there were opportunities to use the data to surface issues and to observe any useful trends that could inform wider practice as well as those taking part (each of whom was sent a report of the findings).

Within the analysis, an early decision was taken not to be unduly concerned about, or distracted by, over-analysis of the data, for example segments within the data such as demographics or organisational issues that appeared. The risk from this could result in insights being missed; however, I kept the focus on the evaluation, rather than wider issues, by ensuring that what was interpreted from the data was linked to the objective and not polluted with a personal need for intellectual curiosity or diffusion.

Content Analysis was again used as the main driver for the analysis; however, some simple statistical analysis of closed-ended questions was carried out to examine internal consistency of the questionnaire. All the factors considered in Phase One were considered and repeated to produce a pragmatic survey result that could surface meaning from current practice. The data will be held for 6 years and then deleted. A more complete discussion of ethical standards and processes is contained in section 3.9.

3.4.8 Phase Two – Conclusion

During the life of the dissertation, my objectives moved from this being a survey of attitudes about Kirkpatrick and the possibility of creating a new paradigm, to the practical creation of a new framework. The two Phases shown above were interesting and informative in themselves, but became superseded in importance by Phase Three, where a new framework was developed and tested as part of this academic process and in tandem with projects as a professional evaluator.

At this stage, it is possible to undervalue the meanings and see them as being somewhat disconnected from the rest of the report; however, I prefer to think that what had gone before served to draw a line under the 'conventional wisdom' and give a range of clues, ideas, and suggestions as to what could come next.

However, as an ethical point, I have spoken to every person who took part in Phase 2, to thank them for their inputs and to inform them of the work that followed their contribution, as I was conscious that they should feel they had added value and could see the 'difference they had made'; from the giving of their time and contribution freely, which was so gratefully received.

3.5 Research Phase Three – Formulation of a New Framework – Overview

Phase	Design	Deployment	Analysis
Phase Three 2014 – 2016	<u>Stage A Framework Creation and Sampling</u>	<u>Stage B cont.. Process Operation</u>	<u>Data Comparison</u>
Stage A – Framework Creation	Framework formulation - Stage One Ideation – reading	Expectations of forecast accuracy collected from Case Study	Third Party results using 'gold standard' process results turned into the 'benchmark'

Stage B - Case Study	and concept formations - Stage Two - Testing of the concept – workshop of 12 experts Creation of initial forecasting accuracy benchmarks	group Retrospective process tried on 27 courses where benchmark information existed	against which the variance of the forecast is measured Cost profile created Outcome comparison
	Stage B Purposive Sampling to select Case Study companies – 7 companies with 6 L&D practitioners	New framework process run in 7 companies and over 94 courses Third party 'gold standard' process then completed in same companies and same courses to create the benchmark score	

3.5.1 Phase Three 'A' – Formulation of a New Framework – Creation Process

I have decided to devote some space at this point to show the formulation and outline of the new framework in order to help situate the decisions and method for the case study segment of the research.

In formulating something new, such as a new framework, two distinctly different processes can be at work. Kirton (1978) defined these as being about either Innovation (making new leaps in thinking), or Adaption (building on an idea that is present). He defined a series of 'conditions' or steps that could be helpful for the process, these also are found in the work of De Bono (1970) and in his ideation processes. This table outlines those stages and how they fit with the research for this project:

Stages	Activity	Research
1	Exposure to a wide range of thinking – either	Literature

	from within the existing fields of knowledge – often popularly called “rolling the grass of new ideas”, that often generates the ‘light bulb’ moments so beloved of creativity thinkers! (The use of specific tools, such as analogical or metaphorical thinking, PMI, or even more structured tool-kits, such as the Six Thinking Hats or Decision Trees, are used to guide and structure thinking)	Review and personal work following reflection
2	‘Delphic influence’ – discussing concepts, socialising ideas and testing frameworks with authorities in the field	Workshop with colleagues
3	Idea refinement and modelling	Planning for the Case Study
4	Plan and implement	Case Study implementation and review

3.5.2 Phase Three ‘A’ – Formulation of a New Framework – Stage 1 – Creation Process – Exposure to new ideas

Detailed reading of the evaluation literature led to new insights and ideas. For example, Spitzer (1984) and Dessinger-Moseley (2006) widened my perspective to see that concepts, perspectives and measurements about evaluation options outside of the usual solutions were possible. And, more encouragingly, these more stimulating ideas were situated in the Decision Theory continuum of the evaluation field. This further reinforced my view that Decision Theory was a useful approach in which to situate evaluation practice and any framework.

My thinking was stimulated by exposure from non-evaluation literature sources began with my fascination with the work of Ferris (2007), whose podcast series interviewed a range of unusual people in his quest to ‘deconstruct success’ and one of the most significant was with Narwal Seti, a successful Silicon Valley investor. A key mantra of his is to read as ‘widely and randomly’ as possible, suggesting that creativity is stimulated from finding diverse perspectives around similar problems. This approach is very different from the more formulaic approaches in knowledge-building within Social Science research, but served as a useful contrast to that reading method deployed in the literature review.

During the podcast series, both Ferris and Seti recommended the book 'Sapiens', by Harari (2004) which is a brief review of the social anthropology of humankind over the last two million years. During one of the chapters, the concept of Jacob Bernoulli's Law of Large numbers was outlined. This is the idea that the larger sample of data becomes, it displays a closer regression to the mean, often codified, in recent literature, as Common Limit Theory.

Whilst covered more thoroughly in the Literature chapter, Bernoulli's idea sits at the heart of many forecasting and predictive tool-kits. However, in order to create a forecast, a clear idea of what is needed to be known has to be decided first. I have always seen evaluation as part of Decision Theory – i.e. you figure out what you want to know and design a process to find it. This is an idea at odds with the prevailing evaluation paradigm that has a single process that finds out whatever anyone decides to contribute to the process. The Law of Large numbers and its application to evaluation forecasting by using metrics and a relevant set of processes suggested a radical option for a new framework.

3.5.3 Phase Three 'A' – Formulation of a New Framework – Creation Process – The 'light bulb' moment

So whilst operating both formal and informal reading processes, my ideas developed, built around the 'light-bulb' moment of applying the Bernoulli concept to the subject of evaluation. My reflection was guided by the conclusions that it is a robust and well-tested method for prediction, with successful and accurate results tested over time. If evaluation practice could harness a robust method such as this, (arguably much more legitimate than the existing evaluation approaches), it could have legitimacy based on its application elsewhere.

As part of the reflection about the Bernoulli concept, the idea of Utility Theory had also still appealed to me as a concept that had still not been fully exploited. Its claims to quantify behavioural change, through the use of Standard Deviation analyses, had initially been dismissed because of insufficient credibility; partly because of insufficient data to drive its application. However, with the rise of big data and the school of thinking around Engagement, linking the concept to forecasting, also suggested another area of innovation that could be developed. In addition to this, the idea of 'Intention' suggested a pragmatic process link between intention and forecasting, as both are based around future-based concepts of Decision and Action Theory, and fit naturally with embedding learning and helping to cement the probability of a successful forecast, through effective Learning Transfer, a process that the literature suggested was increasingly operated by the L&D function.

In synthesising the different areas of thought, as well as day-to-day learning from practice, a number of drivers were created to show the differences in approach between the existing evaluation paradigm and the new framework. The drivers themselves may also stimulate a paradigm shift if they answer the needs that exist in the world of practice and are not answered by the current paradigm.

From	To
Checking on whether any learning impact was achieved	Forecasting the return and ensuring it is delivered
A focus on behaviours in isolation	A focus on behaviour within process
That post-course evaluation is complicated, lengthy and has low rigour due to poor returns	That post-course evaluation simply checks that the factors for value creation are in place and will be triggered
From vague learning objectives	To targeted learning points and actions
That the learner should have fun	That the learner should have 'intent to action'
Scrap learning from 'entertainment'	A move to output focus rather than 'time spent'
Evaluating without purpose	Having an evaluation strategy
Having one process to cover every evaluation need	Deciding what needs to be known and selecting the appropriate process
Being driven around the capabilities of the LMS	Ensuring the LMS delivers what the process demands
Measuring what is measurable	Deciding to measure what is required
Being precise, difficult, and time consuming	Being accurate, straightforward, and resource-light
Transaction costs/process measures are used to reduce budget	Transaction costs/process measures are used to innovate
Allowing the trainer to set the learning agenda	Specific and targeted learning points and actions
Seeing ROI as the goal (the 'holy grail')	Seeing ROI as a secondary or extrapolated goal after organisational measures
Seeing evaluation and metrics as a chore or a threat	Seeing evaluation and metrics as a means for improvement and growth
Seeing evaluation as a unique process and used as a 'stick' after the programme	Seeing evaluation as integrative and part of the transfer process
Measuring subjective outcome estimates	Using salary multiples to ensure an objective number exists within the process

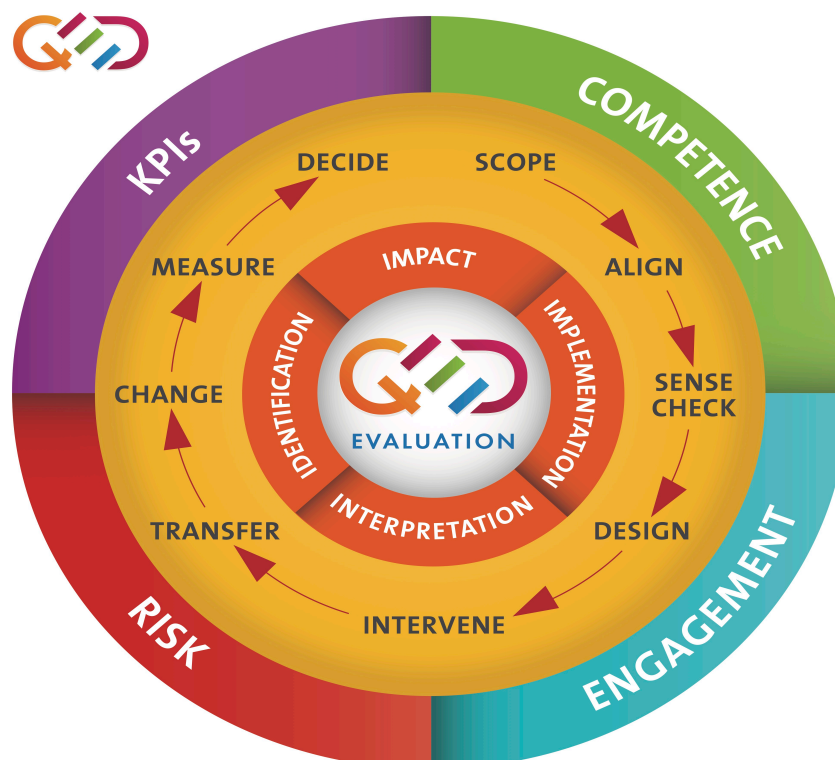
In adapting and synthesising these concepts from the literature (both formal and informal), the set of aspirational drivers had been created against which a new framework could be designed and judged in the

longer term. The framework would need to integrate the drivers and the solutions from the 'light bulb' process, in order to be effective. The actual process of framework-building, as well as process and software-creation, then took place over an extended period of time, before it came to fruition. The actual framework is outlined in the next section.

3.6 Formulation of a New Framework — An Overview of the new Framework — The QED Evaluation Strategy and 4 I's tool-kit

3.6.1 Introduction

I have decided at this point to give an overview of the new framework, in order to help contextualise the discussion around methods etc. This is a pictorial representation of the new framework. It shows the key ideas and processes contained within it and is dissected below to explain what it is and how it works.



Building on the work of Brinkerhoff (1988), Phillips (1999), Anderson (2007) and Passmore (2012) as well as in my own experience, the following table outlines the most identified approaches to evaluation carried out by many organisations and reinforced by findings in both Phases of the research in this document, particularly where Kirkpatrick is the prevailing paradigm in operation. Whilst this is simplified for effect, the approach is resource intensive as the L&D function has to allocate resource from a full workload and traditionally delivers insufficient insight, in effect, simply finding out whether a course 'worked' some time after delivery.

Stage	Actions	Considerations
<i>Approach</i>	Evaluate all courses within the portfolio using a single method based on a post course questionnaire	Level of response rate? Did delegates enjoy the programme?
	Evaluate all courses using a more robust summative process to investigate what has happened	What has changed? Did the course work?
<i>Analysis and Reporting</i>	Analyse data collected through a LMS or a specific survey	Was the response sufficient? Can we report the results? Is anyone interested in the outcome?

New Framework for consideration

This is the breakdown of activity using the new evaluation framework and is geared to identifying what needs to be known in advance of delivery to ensure that the outcomes desired are purposely produced and not left to chance. The framework develops a more strategic approach by considering internal and external stakeholder needs as well as pragmatic internal resource constraints. The move to allocating the majority of

courses to 'resource light' quadrants ensures time can be maximised as required.

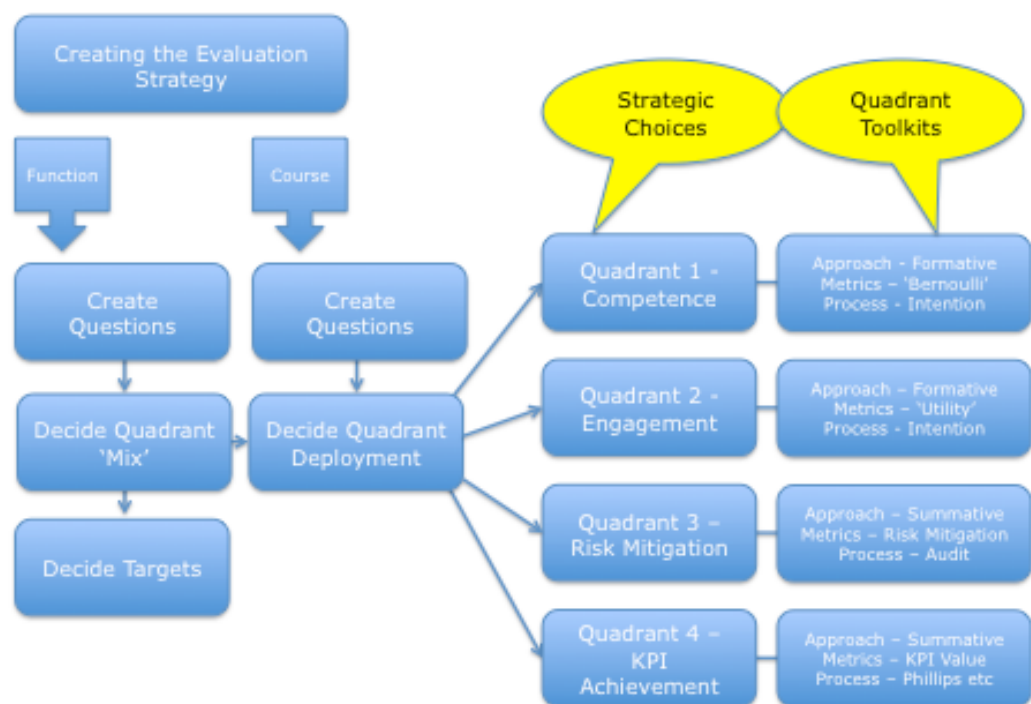
Stage	Actions	Considerations
<i>Create the Evaluation Strategy</i>	Understand the decisions that need to be made post course to drive the appropriate evaluation approach	For example: Should this course continue? Should it be improved? Could it be delivered another way? Is it providing the desired outcomes? Is it effective in terms of time and budget? Could another intervention deliver a better outcome? Should the supplier be replaced?
	Decide the most appropriate method to generate the data to inform the decision. As each quadrant contains toolkits and methods for data capture and analysis, select the appropriate quadrant to best provide the answers sought	Identify which course fits currently into a quadrant – is that appropriate/desired? Should courses sit within more than one quadrant?
	Decide the quadrant 'mix' and create any extra process to ensure that the data can be collected and analysed	Create the learning points and actions that will, with appropriate levels of intention, deliver the data to answer the questions needed to be answered
	Create the appropriate targets for both learning and	Driven by the original questions to be answered

	subsequent outcomes	
	Determine the level of available resource and external requirement	Audit the level of internal resource able to collect and review the data. Understand the actual level of external stakeholder need
<i>Execute the Evaluation Strategy</i>	Having decided the quadrant 'mix' for all courses in the learning portfolio deploy the appropriate quadrant toolkit for each course	This will be driven around needs and available resource to ensure that courses use appropriate process determined by the questions to be answered
	Quadrants One and Two allow for rapid forecasting of results	Ensure that learning intention is sufficient for each course

Each quadrant has within it a series of toolkits, metrics and processes that are deployed to drive the data and analytic requirement as illustrated in the following diagram.



Whilst describing the framework as a series of process steps, the actual L&D function are able to engage with the framework at a number of different points. For example, they may want to develop a strategy, or use a quadrant to formulate questions and process, or simply use a tool-kit. Their access point is driven around their individual needs and the questions they seek to answer. In effect, there are three main aspects to the framework: the creation of an evaluation strategy, developing actions and priorities by deploying the process generated from within a Quadrant, and by blending together the tool-kits within the Quadrant, as shown in the diagram:



3.6.2 Framework Overview – Approach and Operation

In practice, the evaluation framework is a consultancy offering, therefore the L&D practitioner can elect to engage with different elements of the framework, or mix and match elements according to their needs. However, if using the whole framework, the practitioner must first create an overall strategy and approach. As the framework is situated within an

idea of Decision Theory, the strategy must answer questions posed by L&D (or another stakeholder), as part of its approach, some examples of decisions are outlined below.

The starting point in the framework is to decide what is being evaluated; for example, if deciding on an evaluation strategy for the function (or budget), that strategy is based on what information L&D need to discover, in order to allow them to answer relevant question(s) posed by themselves (or another stakeholder). For example: is the budget well spent? Could costs be reduced overall with the same level of training impact? Can the budget flex to cope with a large programme without reducing quality? Could another party, other than the incumbent team, generate better results with that budget? Alternatively, if deciding on an evaluation strategy for a particular programme or course, the evaluation strategy is based on what information L&D need to discover, in order to allow them to answer relevant question(s) posed by themselves (or another stakeholder). For example, does the course need to improve? Do people value the experience? Is there sufficient learning, is it worth the money? Should the trainer be replaced? Does it make a difference to the bottom line?

In building the strategy, suitable 'targets' (or KPIs) can be created, that the actual course (etc.) will need to deliver. These targets usually link to a scorecard approach that many organisations operate. For example, targets could include: financial elements (ROI, Cost per Head etc.); Process elements (workshop actions, e-Learning delivery etc.); Customer elements (satisfaction, repeat purchase etc.), and People elements (learning, wellbeing etc.), depending on whether targeting the function on its internal process operation, or the external aims of the courses for the recipients or stakeholders of the learning.

Having generated the decisions to be answered, the next stage is to link the needs generated by the questions, to the solutions contained within four broad quadrants that reflect different approaches. Quadrants One and Two are very resource-light and rely on formative evaluation,

Quadrants Three and Four are based on summative evaluation methods, and answer questions that need to examine longer-term impacts. In this way, an 'evaluation mix' can be created, driven by needs as well as available resources. Within each of the quadrants is a series of tool-kits (for example, forecasting software, questions for delegates and managers, metrics, training to help managers with transfer) that allow the practitioner to ensure that the data is created and collected to answer the questions and achieve the targets previously created.

The criteria used by the L&D practitioner is to use the quadrant that represents the most cost-effective approach to deliver the information they need to produce the evaluation outcomes they require. An overview of the quadrants follows; however, Quadrant One is explored in greater depth below this section, as this is the area that the research within this project will determine to be accurate and legitimate for practitioners, through the use of a Case Study approach.

Quadrant One evaluates the development of 'Competence' and the majority of training delivered should be evaluated using the tool-kits contained in this quadrant, as it is both accurate, cost effective and resource-light. The tool-kits use a Forecasting approach and measurement metrics, based on the Law of Large Numbers principle which are refined, using a range of specific learning-drivers, to give a precise forecast. The L&D internal process change for this quadrant is based on the levels of intention measured at the end of the course, and intention scores are used to ascertain the likelihood of learning-transfer and generate a probability score for whether the learning will be deployed in the workplace – this allows the L&D team to decide whether to intervene to affect the outcome or not (for example, if the likelihood of the learning being transferred is low). More detail follows later regarding Quadrant One.

Quadrant Two evaluates the development of 'Engagement' in the delegate. It is likely that this quadrant will be used less frequently than Quadrant One, usually for 'difficult to measure' courses, or in conjunction

with another Quadrant. The term Engagement is used here to describe 'soft' factors that drive behaviour, for example - levels of emotion, morale, mindset, motivation, confidence, and commitment to using the training content.

The tool-kits within this quadrant are, again, based on Forecasting. However, this time the measurement metric driving the result is based on Utility Theory which forecasts an output, based on standard deviations of behaviour. The Utility Theory allows for a salary allocation to the forecast levels of behaviour change. The previous levels of dissatisfaction with Utility Theory, outlined in the literature review, means that this quadrant may be initially used sparingly, but is useful for courses that concentrate on very soft skills or conceptual personal development, which do engender positivity, but do not always exhibit an immediate, tangible workplace impact. The L&D internal process change for this quadrant is based on the levels of intention measured at the end of the course. The intention scores operate as described in Quadrant One. Whilst this process is also new and innovative, I decided, for the purposes of the Case Study, to treat this stage as being out of scope, in order to focus on the impact of Quadrant One. Some of the implications of this may affect the research outputs, and will be discussed within the research findings chapter.

Quadrant Three evaluates the 'mitigation of risk' particularly with regard to Health and Safety or other mandatory types of training and is measured using a summative process – the organisational context will drive the importance of this Quadrant; for example, financial services and logistics or manufacturing have a greater need for this type of information, as their mandatory processes are often linked to punitive action (for example, closure of a site), and/or legislative penalties (for example, organisational fines). The measurement metric is a more traditional quantification of risk-mitigation approach. The L&D process would still involve the use of 'Intention' and can give an indication of the probability of impact on the quantification value of risk-mitigation. These risk-mitigation values exist widely (and readily), in the sectors where

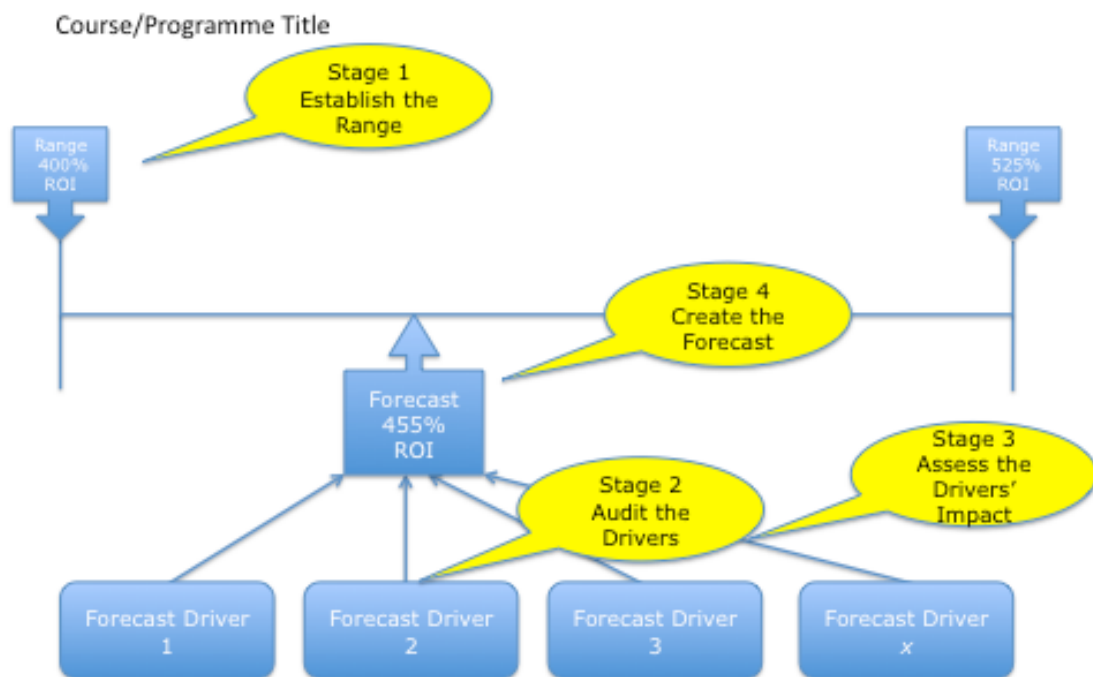
they matter, as they are used frequently for process improvement, regulation, and operational effectiveness.

Quadrant Four uses a more traditional evaluation approach and reflects the occasional need for resource-hungry evaluation, which evaluates the post-course impact of actions positively to affect the achievement of organisation KPIs. This quadrant uses a summative process – the minority of training should sit in this quadrant, because of the complexity of measurement; however, identifying specific courses and programmes that require this level of resource, analysis, and process effort, and being able to allocate that required resource is more realistic if the practitioner is using the other Quadrants effectively. The measurement metric is based on the quantification of impact achievement, having determined the value of the KPI change and the level of impact created by the training intervention. The L&D process used is the more traditional tool-kit of summative evaluation, perhaps using Phillips or Success Case Methods to generate information.

The idea of the framework is to focus more of the evaluation activity in Quadrants One and Two, to leave more time for the more resource-intensive Quadrants Three and Four, which need to be used only if there is a genuine decision to be made that needs that level of detail. In many cases, the data from Quadrants One and Two will help with any decision that needs a more robust summative process to be conducted. However, whilst all of the quadrants have their own approaches and tool-kits, they can all generate a ROI figure, if that is required as part of the decision, and it is my contention that much of the evaluation insight from Quadrants Three & Four can be generated as effectively within Quadrants One or Two.

3.6.3 Quadrant One – The Subject of the Case Study Research

Having outlined the overall framework, a further discussion of Quadrant One is appropriate as it is the subject of the research in this document. An overview of the Quadrant tool-kit is shown here:



To explain the process and tool-kits, a programme or course is first selected by the L&D team. A range of forecast ROI is generated from all the data known about the broad type of courses. This range builds on the 'Bernoulli' idea, (Law of Large Numbers) where forecasts revert to a mean when being able to use sufficient quantities of data to decide what that mean may be. As I now utilise my own proprietary software, in effect, the forecasts are generated from the data used from thousands of previous evaluation projects as well as the data from other LMS providers and external data sources, who collaborated with my organisation to build a shared benchmark database. One of the benefits of this approach is that, if the forecast falls outside of the range having conducted the forecast assessment, there can be a discussion whether a course or programme should actually run, or be amended to help it generate the returns required by the forecast.

However, when perfect information is not available, (large enough data of a specific type), a range is created and a final figure is created from the application of factors that drive greater precision, for example, superior learning practice and learning transfer. Those factors are audited to

examine whether they exist and, if they do, whether they are operated at an acceptable level of competence within the specific L&D function.

An example of the factors may include:

Stage	Factors
<i>Commissioning</i>	Commissioning is robust with proper challenge of 'need' as well as expected outcomes, process changes and targets Delivers a 'return on targets' concept Selection of a trainer prepared to deliver against targets An appropriate evaluation process is chosen
<i>Design</i>	Specific learning points are developed for each phase of the programme and against benchmarks for the type of course Skill levels 'from – to' are created A range of actions is created – each is valued
<i>Delivery</i>	Focused on the core elements to drive high intention. Immediate assessment checks intention elements and targeted actions Assessment of learning achievement Increased use of learning application methods rather than theoretical knowledge
<i>Embedding</i>	Learning is supported by: <ul style="list-style-type: none"> • Specific management action • Linkage to other HR process • The level of 'learning pollution' in the culture • L&D resource • Blended support • Interactive learning support • Implementation of actions
<i>Evaluation Process</i>	Immediate review of intentions and actions achieved by trainer and learner Optional post course audit/sample of: How many actions were implemented Whether those actions delivered the forecast returns

This selection of factors is an example of part of the mix of elements audited to ascertain the accuracy of the range. Clearly, the value created from this process allows the L&D practitioner to begin to continuously improve their practice through the use of the evaluation metrics. Using this approach, each programme starts with a range, and the assessment

of the drivers creates the final number. It is the accuracy of this overall process that will be assessed in the Case Study.

Quadrant One has many areas of difference from the traditional Kirkpatrick approach.

Approach	Areas of Difference
<i>Process</i>	<p>It is focused on 'intention to execute' and learning transfer measures rather than the four levels in the Kirkpatrick framework and is unlike anything in evaluation practice other than the Basareb (2007) framework.</p> <p>L&D generate the actions that the learner need to implement to drive the impact that the forecast identifies. L&D carry out a quality audit approach to ensure that the course generates the actions required that ensure the impacts drive the correct forecast.</p> <p>L&D can affect the forecast by identifying people with low intention and affecting further activity e.g. advising managers (etc.).</p>
<i>Measurement</i>	<p>The idea of measurement of reaction, or behaviour, is removed in favour of intention and process change.</p> <p>The measurement process is based on the forecasting of a result using the Bernoulli concept rather than any form of summative review.</p> <p>The initial range is established from large data samples of similar courses.</p> <p>Specific Drivers are identified that move the forecast within the range from (e.g.) High to Low dependent on whether they exist.</p>
<i>Outcome</i>	<p>Kirkpatrick considered Learning Objectives to help consider Return on Expectations. This is replaced.</p> <p>The results are driven by a 'target' score that drives the achievement of the overall L&D budget – this target score is linked to the decision that needs to be answered – for example, value for money or sufficient learning etc.</p>

The research is therefore focused in determining whether the idea that underpins this concept is legitimate, robust and accurate enough for a practitioner base through the use of knowledge that is applied and tested through results, as suggested by Russell (1912).

3.6.4 Summary of the New Framework

The new framework creates a strategic choice between a quick and accurate forecast or a slow and purposeful summative evaluation. The decision in hand generates the strategic choice. Therefore, the new framework the features of Decision Theory, particularly as suggested by Hubbard (2007). The research will allow me to examine the accuracy of the forecasting method against the benchmarks created by the current 'gold standard' approach, using the Phillips method to calculate ROI. This will create a realistic test of the approach and see whether the reductions in the cost of the forecasting approach would be a worthwhile saving against the expected loss of accuracy in using forecasting as the primary method of calculating value.

Having deviated to explain the framework, the following sections continue the discussion around methods.

3.7 Phase Three 'A' — Formulation of a New Framework — Creation Process — Stage 2 — Testing using a panel of experts

Having outlined the research framework above, this chapter focuses again on the idea creation process. As part of innovation practice, the idea of Delphic challenge can be a useful stage within the creation of a new idea. I again operated a form of purposive sampling in order to gather together a small group of people who could be relied upon to meet a number of criteria, including:

- Keeping the information confidential
- Having extensive experience in evaluation and commercial activity within L&D (as trainers or managers within an organisation or as trainers or Directors of training companies)
- Being challenging and commercially realistic

The group of colleague evaluators would work together to discuss the concept, identify the weaknesses and challenges for the framework and to identify the commercial challenges that could exist.

The group consisted of five senior evaluators, five associates (freelance evaluators) and two colleagues working in an evaluation organisation not currently involved in the L&D sector but with extensive experience of the sector. The use of 3rd parties in Academic Research is discussed further in Chapter 3.9.

I recognise a number of weaknesses in this process:

Weakness	Action taken to mitigate the risk
My personal bias in selecting people with whom I have a continuing relationship	I explained the context of the meeting and my expectations of people. Ground rules were established (for all of us) and guarantees made about the need for impartiality in the process. A further discussion of an ethical stance is covered in 3.9
My sample size was limited to people I knew and could trust	I had little choice with this, but ensured that the people selected were robust contributors; those with views and opinions I respected.

Their views could be watered down in case I did not want to hear a difficult message	I engaged a facilitator to run the meeting process and to ensure all meeting outputs were captured – irrespective of the content of the message
The process would be open to my interference and hi-jacking	The facilitator designed a strong process of plenary and small group sessions to ensure that conversations would be free from any distraction and undue influence from my input (the outline of the workshop is contained in Appendix Five)
The group of people may be unduly disposed to innovation and fresh ideas	Each person was briefed to articulate their own views and opinions. Many of the group were considered highly data rational and could be influenced only if the statistical and mathematical logic stood up to their scrutiny. Some of these people are noted cynics and risk averse and are often regarded to be resistant to change in the community
The group could be unduly helpful in order to support me in my research	Whilst some of these people are colleagues, none had any vested, financial or personal interest in the outcome and are far from romantic when it comes to making commercial, technological and financial decisions. The facilitator also acted as an external brake on undue optimism or helpfulness

Whilst broadly positive and optimistic, they left me with five key challenges that they thought would be solvable but which may make the framework difficult to be easily accepted by the wider L&D practitioner base.

Challenge 1 - Surely all courses are different?

Having watched, trained, delivered, assessed and evaluated thousands of courses it is fair to say that most learning has more in common than differences. Most companies:

- use similar course learning objectives (often weak and uninformed)
- have similar course outcome measures (often meaningless)

- use similar conceptual frameworks within the learning
- use trainers who have learned similar methods and techniques and use similar learning techniques and processes (discussions, disputations, skill practice, coaching etc.) within the learning environment
- rate trainers based on popularity rather than effectiveness
- use training methods and techniques that have already proven to be popular to previous learners

Challenge 2 - Surely all learners are different?

This is also true, but it is possible to calculate three vital areas that affect the majority of learners:

- Their personal motivation to implement the learning
- The process within the workplace to measure their performance or 'contribution'
- The culture in which to be able to transfer the learning

This means that the L&D function can calculate the desired productivity returns, and have a dialogue with the organisation, to ensure they are achieved. It also shows the point at which the 'hand-over' for learning takes place between the L&D function and the Manager and Learner.

And again, courses themselves suggest that learners are less individual than originally thought – a Health & Safety course across sectors, organisations etc. has a similar implementation rate, irrespective of the type, gender, status, role, or style of the learner.

Challenge 3 - Surely the measures are as flimsy as ROI?

The ability to release the value of training by using payroll, as well as the theory of large numbers to predict an outcome, links two strong processes with a fixed point of data. This is far more secure than the prevailing paradigm

Challenge 4 - Surely we want to know ROI?

This is sometimes true and this framework also contains processes that generate that metric. However, it is also the case that we need to carry out post-event evaluation *but only where and as required* – this means that the process will have to fit the needs of the decision, not the other way around, and that its legitimacy is greater, as the correct process is tied up with the decision required. Also, as many people use 'ROI' as a shorthand term for a range of meanings, the framework challenges what they actually desire when it is requested as an output.

Challenge 5 – Surely forecasting is no more than guesswork?

Most statisticians understand a well-governed stochastic process, built on strong logics and mathematical principles, has an excellent track record of success, assuming the length of a forecast is within sensible temporal boundaries. In addition, it is possible to create pragmatic outcome measures, based on existing ideas that can multiply the returns from this new approach.

3.7.1 Phase Three 'A' Group Benchmark Challenge

One of the subjects discussed with the group was the degree to which confidence would be built if the accuracy of the forecast was sufficiently accurate. This accuracy could be further reinforced by the cost savings; however, to be legitimate enough for the L&D practitioner to believe in the result and defend the result internally, it should stand on its own.

Therefore the following question was asked of the group:

- What level of forecasting accuracy (plus or minus), would persuade you that the level of forecasting would be valid?

Each answer was collected and, in the remaining testing phases, this data was used as a benchmark guide. Interestingly, ten of the 12 evaluation experts believed that the forecast should never exceed the benchmark in order to avoid over-optimistic forecasts (the problem that had negatively

affected the original credibility of Utility Theory). The answers are contained in Chapter 4.

3.7.2 Phase Three 'A' – Creation of New Framework – Summary

Kirton (2003) and De Bono (1970) both stress that innovation is surprisingly rare and my subsequent analysis of the framework shows clear adaptive elements from complementary areas of work and research, as well as clear debts to the formative school of evaluation thinkers including Spitzer (2005) and others. However, within the field of evaluation, there was enough innovation within the process, tool-kits, and method to create significant interest in the evaluation practitioner base when launched at the CIPD conference of 2016. A series of short workshops was presented, as well as a fringe workshop to outline the new process and, as a result, over forty significant client conversations have ensued as a result.

However, at this stage, the idea of the Case Study is still simply to test whether this new Forecasting process would work as effectively as the 'gold standard' version of evaluation, and the thinking and planning began in earnest. In other words, 'will it work', not 'how does it work'!

3.8 Research Phase Three 'B' – Case Study – Considerations

The Case Study had two main thrusts:

- To examine the legitimacy of using a forecast within the framework by comparing the results obtained using the new process with those generated by an independent third party carrying out a 'gold-standard' summative approach
- To examine whether the new approach would be cost effective by comparing the cost and resource implications between the two methods

When first constructing the methodology for the DProf process, I wavered between this being structured around Action Research, as opposed to using a Case Study. One of the many attractions of using a Case Study is that Gray (2009) reinforces the idea of Yin (2003), that they are used when a causal relationship or proof is required, rather than a simple description, and that blending qualitative and quantitative methods is also acceptable, as both an inductive or deductive approach can be used. Also, a number of different data points can be considered and generated in the process, as the Case Study is being used to investigate a phenomenon with a 'moving' or real context.

This somewhat 'messy' situation is one that, as a consultant, I am used to dealing with on a daily basis, and to which I have ethical and process guidance to help deal with the complexities of culture and context.

Whilst Gray (2009) points out that Case Studies are not universally accepted as 'reliable, objective and legitimate', both Yin (2003), and Gummesson (2000), have developed justifications around the multiplicity of examples and cases that can be generated by using case studies, to be able to generalise, as is done in other scientific research methods.

Heeding their warnings that Case Studies can generate huge amounts of data and paperwork, I have focused very clearly on the data needed for the purposes in the research, and resisted the urge to become unduly distracted by other data around culture and context.

Gray (2009) and Yin (2003), suggest a method to develop an appropriate process for the case study method, therefore, the following approach and processes were used:

Stage	Tasks
<i>Define and Design</i>	<ul style="list-style-type: none"> • Develop Theory - Create a framework to forecast the returns from learning interventions • Agree Case Study Companies and agree process and ethics for each • Agree learning interventions

	<ul style="list-style-type: none"> • Prepare all parties for the deployment of the relevant stages in each of the steps
<i>Prepare, collect and analyse</i>	<ul style="list-style-type: none"> • Carry out each case study (all run in advance of the learning intervention) • Individual reports completed and stored where they could not be changed • Learning Intervention took place • Third party evaluator ran the summative process following the learning intervention • Their reports written and collated
<i>Analyse and Conclude</i>	<ul style="list-style-type: none"> • Cross case conclusions drawn and documented using evaluation approaches • Time and resource calculations completed and compared • Actual vs. forecast results compared • Phase Four research carried out

Yin (2003), suggests there are six main sources of data collection for case studies:

Source of Evidence	Approach
<i>Documentation</i>	This was the main method used in the Case Study as it is deemed to be objective and, as collected by myself as well as a third party, less likely to be cluttered by organisational detritus or individual bias, as the collectors of evaluation data work to exacting standards
<i>Archival Records</i>	Not required
<i>Interviews</i>	These were used particularly as part of the summative third party process and not our own. They adhered to their own internal rules and guidelines for this exercise. Interviews were also used as part of Phase Four
<i>Direct Observation</i>	Not required
<i>Participant Observation</i>	Not required
<i>Physical Artefacts</i>	These were used as part of the summative third party process and not my own.

One of the challenges with the case study alluded to earlier was validity and reliability. To deal with this, the following strategies were deployed:

Area of Concern	Strategies
<i>Construct Validity</i> - because of the difficulty of defining the constructs being investigated, Yin (2003)	Each party to the process was briefed around scope, responsibility and role. As I deployed standard, well known processes, as laid down and understood by professionally trained evaluators, this was deemed not to be a problem
<i>Internal Validity</i>	No inferences other than what were required for the project were drawn. In fact, one of the frustrations in the project was not being able to 'fix' problems we encountered that may be part of a consultative approach
<i>External Validity</i>	The case study organisations were of varying sizes and contexts and had been selected to ensure certain types of organisation were not included, so that any generalised claims would not be claimed within a non-existing segment. Any claims at this stage would be 'probabilistic', to reflect the advice given by Lieberman (2000), regarding small sample sizes of cases.
<i>Reliability – This, for me was the key test for the wider practitioner base</i>	Bryman (2007) cites the use of an independent researcher to validate findings, and this process is used within the research. They carried out their own (unseen), research and delivered a report according to their independent findings

3.8.1 Phase Three 'B' – Analysis

In the event, the actual level of analysis is very simple; the acid test was simply 'did the framework work', against the parameters set by the practitioners in the Case Study and, therefore, consists of straightforward challenges:

- Did the forecast outcome perform within the defined parameters of the 'gold standard' outcome, and whether any degrees of variance, were acceptable to the researcher and client
- Did the difference in the process and resource implications of each method, including factors such as tangible cost (fees, time), intangible cost (disruption, administration) create a significant advantage for the new method. Opportunity cost was disregarded for the purposes of this project

One of the challenges of the research was the extent to which I was dependent on a binary result (did it, or did it not work). The Case Study itself was therefore to be supported by the Phase Four feedback, which was put in place to diagnose any problems that would have been generated by a failure of the framework.

3.8.2 Phase Three 'B' Case Study Method — Deployment

The final group that volunteered to take part in the testing of the new method consisted of seven companies, but with only six L&D clients. Some of them had volunteered themselves from the research group above, plus one extra client we had acquired very recently because of their urgent need to develop an evaluation tool-kit very quickly to save their budget/function.

In selecting the research group from a number of volunteers, there are risks from a range of vested interests colliding, to want to create a positive result. However, whilst the entire group had goodwill toward the project, I had ensured that these organisations had a challenging L&D Manager at the core who could (and would), give honest and objective feedback during the process. Two smaller organisations were represented by the same person who was head of L&D for both organisations. This gave us the opportunity to examine the data where the L&D person would not be a variable within the process.

In order to assist with the understanding of each of the case study organisations, I have prepared short pen portraits of each organisation:

Company One:

An international pharmaceuticals' company based in England that created and distributed a range of drugs around the world. They had grown rapidly through merger and acquisition, and were owned by a Private Equity firm at the time of this research. This created a strong culture with

an entrepreneurial, pragmatic approach, which focused on speed rather than precision. They had 1500 people in 14 countries and a small HR team of six spread over two territories. They had a very flexible LMS that allowed them to run highly bespoke evaluation, although the culture rarely had time to regroup and learn – being onto the 'next thing'. The Head of HR was the contact link with our project. Whilst operating evaluation processes, they were resistant to increased evaluation and freely admitted to being bored by it – preferring to focus more on the design and delivery aspects of training.

Company Two:

A UK Housing Association based in England. In common with all housing associations at the time of this research, they were facing a tough economic future because of the changes in funding from government having to reduce costs on an annual basis, whilst delivering improved results. Training was important in delivering mandatory skills, without which the association would have been fined heavily, or be closed. A more traditional culture than usual existed in this organisation, although rapid change was in progress through the hiring of a new senior leadership team. There was a large HR team for the staff of approximately 800, but a small and very ambitious L&D team. L&D wanted to prove their value, protect and grow their budget and grow their reach and credibility. There was a constant drive to merge, divest and acquire housing associations, and there was a strong motivator in evaluation being able to assist with functional and job security.

Company Three:

A major UK Retail Group based in England. With over 6,000 employees there was a major programme of Head Office as well as store-based training. The retail challenge is about the speed of operation and excellence in a wide range of facets; including design, procurement, logistic and retailing itself. There was a small L&D team that delivered the majority of non-store based learning. A highly blended approach allowed for retail staff to have a balance of online and store-based development.

Aligning the fast moving operation with the need for strategic acumen called for excellent leadership development, strong values, and a supportive and challenging culture. Recent cost constraints had seen the L&D budget reduced significantly, precisely because no tangible value could be evidenced.

Company Four:

A European Finance group headquartered in England. This organisation operated in the area of consumer and retail financing, within a franchised structure. Significant development took place with a heavy regulatory burden that had to be taken into account. An sector that had suffered from a poor reputation and accusations of sharp practice, this brand had an enviable reputation amongst its peers. This small OD and L&D team balanced the demands from both the head office function and the franchise network. A suite of offerings was provided that was funded via a levy; this sometimes resulted in problems with learner attitude and embedding. A European parent created different cultural challenges for L&D with a dotted line to the UK organisation.

Company Five and Five 'a':

This 'organisation' is an amalgam of two different not-for-profit organisations. As they were reasonably distinctive, it is hard not to give away their identity. However, they were of a similar size and with similar problems. One was a regulator and the other a professional membership organisation within the same sector. Both of them needed to evaluate for a variety of reasons, including justification in the creation of value from stakeholder funds, as well as a genuine desire to continuously improve their offering. In addition, trustees and external regulation focused efforts on excellence, as the impact of training was an aspect of their external assessment criteria. They operated with well-resourced OD and L&D teams and tended to deliver internally across the HQ and operational aspects of their operation. In both cases they tended to focus their evaluation efforts on the leadership and management aspects of development.

Company Six:

This was a division of an international technology company, based in England. Working across Europe, they delivered applications, support and service to a range of internal customers. The division had some 7000 employees and was part of a greater whole of some 40,000 employees. There was an aggressive sales process, that drove strong growth in their crowded sector, and the rest of the organisation had to cope with the strong sales 'bow wave'. L&D was involved with all areas apart from the sales engine. This meant it could be difficult to prove value, and budget was regularly appropriated away from development, other than for sales. The L&D team delivered all of the training themselves, with little budget for anything other than workshop-based training. The need to justify their existence was constant and they had recently survived a move to outsource the function.

So, in each case, the sponsor was the most senior HR/OD or L&D Director/Manager. Each had their own vested interests in proving their value as part of their role, and were without the time or resource, in most cases, to carry out onerous evaluation processes. As part of this process, and because of the highly sensitive nature of the information being shared, we willingly signed a standard non-disclosure agreement with each of them. Whilst this does restrict the ability to share specific names and organisations, it also creates a strong ethical stance and mutual protection.

3.8.3 Phase Three 'B' – Approach, Process

As someone within the interpretivist paradigm using a phenomenologist approach to construct sense of the different meanings in emergent phenomena, I faced a dilemma when thinking about the case study and the statistical evidence I needed to analyse from emergent data. Hubbard (2014), defines an experiment as 'any phenomena deliberately created for the purpose of observation' from which meaning can be drawn and, through reflection, truths constructed.

Hubbard (2014) discusses the pragmatic approaches needed to prove business decisions as opposed to the methods that often exist in the academic literature, stressing the difference in understanding as a practitioner, whether the outputs add value, rather than passing a statistical hurdle that may or may not add value to the decision in hand. What matters is the creation and meanings in the phenomena with appropriate statistics that meet the needs of the parties concerned. Therefore, in effect, the group of companies reflect a t-sample of an expert data group, where a small number of results can illustrate the ranges within a wider population (without inferring generalisability). This means that any statistics would need to reflect the view of the group and not treat them as a control group, or bring the need to use further control groups. The need for significance, regression modelling etc. is also not required, as testing the views of this group against a wider population is not taking place. Using the phenomenist approach, the group becomes the total population, as we are not proving a correlation with the wider group.

I have also been influenced in excluding processes, such as significance and regression etc., because of the use of forecasting as the core measurement process and find some of the arguments of Meehl (1978), that significance is 'one of the worst things to happen in psychology' and where probability (or forecasting), is considered that Deming (1998) says that 'significance, or lack of it, provides no degree of belief...about prediction of performance'. In addition Jeffreys (1961) reinforces the point that using a p-value is not a valid approach in forecasting. Whilst an academic approach may support a wider use of the null hypothesis, statistical significance and regression etc. 'are not required for a management decision'. Hubbard (2014), also contends that methods such as Bayesian forecasting need not be considered when the group defines for itself the levels of performance required from any probability estimate. However, he recommends the use of Bayes in future research, if questions such as the applicability of the framework were to be justified, or to rebut some of the objections and probabilities of assessing

the further use of the data. That process is beyond the scope of this research.

In creating the parameters for the research within the Case Study, each volunteer company was invited to identify courses or programmes where value needed to be measured. The courses could include any of the following broad criteria that I also wanted to evaluate as well:

- A specific large (or important) training course or programme
- Blended or workshops
- Skills, personal development, mandatory or professional development
- A number of small training programmes
- An entire budget

Each organisation was invited to select whatever it wanted – dependent on its own needs, as well as my requirement to ensure that each element shown above was covered more than once. Because of the range of sectors, needs, and levels of engagement, we were able to establish the following activities to allow us to test the framework, across a range of different events, including:

Organisation	Course	Description of Work
1,5,6	A	Valuation of a whole budget
2	B	Valuation of 30 fundamental job skills programmes
3	C	Valuation of a number (4) of Leadership & Management programmes piloted within a division having previously been run in other divisions
5	D	Valuation of an 8 module soft skills/behavioural personal development programme
2,4	E	Valuation of mandatory rolling cycle of programmes using blended learning
2,6	F	Valuation of series of technical programmes each one day long
3,4	G	Valuation of mandatory health and safety programmes

1,4	H	Valuation of induction programmes
1,6	I	Valuation of an accredited Leadership programme
3	J	Valuation of programmes to third party stakeholders
5	K	Valuation of a rolling programme of CPD
1	L	Valuation of large organisational initiative (e.g. Customer Service KPI change)

3.8.4 Phase Three 'B' – Validity Expectations

One of the ethical considerations for those taking part in the Case Study was there was to be no publication of the actual ROI figures from the courses that were being delivered. The investigation was of the accuracy of the forecast, rather than the actual comparison of ROI. This was important, as some of the ROI figures for similar courses were very different, and it was important not to confuse the judgement about how and why the ROI was at a certain level, with whether the forecast could deal with this. It is also beyond the scope of this project what impact the various criteria had in creating the forecast, as that had been developed using a different process. What mattered was whether this approach, built on and utilising the concepts of Decision Theory and forecasting, could deliver against the benchmarks set from the most respected, current evaluation measurement method.

In order to create a method to consider whether the results from the case study could be considered to be 'useful', and meet the ethical commitments, I decided to consider some of the tests laid out in a process recommended by Armstrong (2001) in his evaluation of the most useful forecasting methods, to ensure that good practice was informing my approach. One test included the requirement to 'look back' and compare previously used data for accuracy, as well as the use of an independent third party to create an assessment against which to judge the forecast method and results. As well as understanding any personal sources of bias, the key method was to compare outcomes obtained by

different measures, and this is the primary method deployed in this study.

3.8.5 Phase Three 'B' – Measurement Criteria

The 'gold-standard' result that would be generated by the Jack Phillips process (for example, Training Course 5 generated an ROI number of 350 per cent) would be turned into a score of 100. A simple percentage variance from that result, using the forecasting method, can be generated. To illustrate, if our score of the same Training course 5 generated an ROI number of 315 per cent ROI, then we would be at 90 per cent accuracy – but under forecast. Similarly, if we generated an ROI score of 385, we would be still be at 90 per cent accuracy - but over forecast.

What mattered was the actual level of expected forecast accuracy that would persuade a practitioner that the results were credible and sufficiently accurate to generate confidence in the process. As a result of the work of the 'expert-group', a decision had been made to avoid any over forecast, in order to avoid undue optimism in the framework.

Whilst I had polled the expectations for forecast accuracy in the expert group earlier in the process, I decided to understand the real expectations and parameters from the people actually experiencing the process; I decided that the most appropriate method for the Case Study would be to ask the client L&D practitioners. This also negated the need for any Bayesian analysis, as the sample would then define the performance expectation required from the framework. This met the test outlined by Hubbard (2014). Therefore each person was asked the following question:

- What level of forecasting accuracy (plus or minus), would persuade you that the level of forecasting would be valid?

Each answer was collected, and the results would be reported against their expectations, which are shown in Chapter 4. All of the practitioners believed that the forecast should never exceed the benchmark. This reinforced the views of the expert group as well.

3.8.6 Phase Three 'B' – Case Study Deployment – Data Capture

Having established the programmes to be run, a number of processes had to be operated to allow the case study research to take place:

Process One – Looking back

As the new method involved forecasting, a decision was taken to abide by the traditional tests espoused by Armstrong (2001), to examine assumptions on a sample, looking backwards to see whether the forecasts would deliver against known results. Within my own practice, we had access to existing data from evaluations carried out by colleagues within external clients using a summative method and for which I had no knowledge of the results. To ensure ethical compliance, I contacted each of the clients and, having gained consent, I then operated the new process as far as possible to see whether the result would be possible with only partial access to all the required information. The results are shown in Chapter 4.

I was able to operate the process on the following courses:

- A 2 day leadership programme for middle managers in an automotive environment
- An executive coaching programme (8 sessions per person) for Directors in a media environment
- 20, one day Skills workshops for managers and non-managers in a charity. Subjects included for example: Presentation Skills, Dealing with Difficult People, Having Difficult Conversations, Stress Management, Introduction to Process Improvement etc.

- 5 Mandatory courses (half day) in a medical environment including, Manual Handling, Corrosive Materials, Back Protection etc.

The date from the summative process was adapted to become a benchmark number of 100, and the forecast method then was compared as a percentage difference from the benchmark, and shown in Chapter4.

3.8.7 Phase Three 'B' – Main Case Study

3.8.7.1 Forecasting approach

For each of the programmes/courses selected by the Case Study participants, the forecasting approach was operated, involving desk work with the organisation and the L&D practitioner/client, to understand current practice in the following areas:

- A detailed review of learning points, and learning transfer processes
- A detailed idea of the learner profile to establish motives and timing etc.
- An average salary value calculated for delegates plus a calculation of tangible and intangible costs
- A review of core L&D process to establish internal process for data capture
- A review of trainer motives and approach

This allowed the creation of the forecast result, at the appropriate point within the forecast range, determined by the original Bernoulli concept, that is based on the huge amount of data currently held around the same types of programmes. As this process is formative, it runs before the courses take place, or any suitable point in the deployment of a programmes, as the result does not depend on the learner. In addition to this usual process, a log of time was kept per course and company to be able to compare the time commitments between the processes.

3.8.7.2 Benchmark approach

A full 'Jack Phillips' (Gold standard') methodology was run on each of the courses by an evaluator within a third party organisation. This third party evaluation was funded totally by me as part of the research for this document. The company used is a highly regarded evaluation practice that had held merger discussions with us; therefore we had full knowledge of each other and a signed Non Disclosure was in place at that time. As the Phillips' approach is summative, the courses needed to be run and sufficient elapsed time to have elapsed before their process was complete and could be compared to the forecasts.

The outline of their stages are as follows:

- Understanding and profiling of learning outcomes
- Course cost profile created, based on tangible, intangible, and opportunity costs – this is used as part of the cost/benefit number, that subsequently forms the ROI number.
- Initial reaction at the end of the course
- Follow up of delegates at one week and one month, following the close of the course or final part of a programme, to determine impact. Face to face interviews or online surveys are used to consult with the delegates and the managers or peers of the delegates, depending on the course
- Separate impact review interviews if insufficient data had been collected

In addition to this usual process, a log of time was kept per course by the third party evaluator. Fees for the client company were simple to calculate, as they were charged to me at a commercial rate. Each process generated its own data outputs, that were collected using software systems. In order to make sense of the third party data for a non-commercial purpose, each set of data was converted into outputs that could be charted using an excel spreadsheet. This allowed charts to be created of a similar style so that the host company, in analysing its

results and outcomes, would not be swayed by superior or less glitzy reporting.

In setting up the project with each of the case study organisations, time was spent with each Head of L&D, to discuss with them the new framework and the requirement for development for each. In every case, development was only given to the L&D team to help them understand the process and the extent to which the criteria for making judgements was being applied in their organisation. In sharing the criteria, the L&D team could choose to build an action plan positively to affect the impact of courses, when our process had been completed, if they so desired. In this way, the case study organisation gained a tangible benefit for taking part in the research. It was tempting at this stage to develop each of the L&D teams in the whole methodology, and attempt to evaluate that outcome but, in the event I focused on the aims of the research – the part that encapsulated the heart of the new framework, proving that forecasting could generate accurate results in a cost effective manner.

3.8.7.3 Case Study Data Analysis

The same process discussed earlier in the 'looking back' process, was used to produce the outcome. The Phillips' 'gold standard' number was turned into a score of 100 and the forecast result could then show the percentage variation, without showing the actual ROI number, as those figures are commercially sensitive. Actual numbers are used without use of significance or regression as the proof of legitimacy had to meet a commercial standard, rather than an academic standard, as it was only these numbers in which the L&D teams were interested. In the event, the data was simple to create in Excel and presented in a series of straightforward charts in Chapter 4.

3.8.8 Data storage and confidentiality protection

The commercial sensitivities in this type of commercial evaluation research go far beyond the rigour and nature of confidentiality of a

research project. All the data is stored in commercial databases for the purposes of this report. However, due to the nature of the NDA, the specific attribution of outputs to named organisations is prohibited. As part of the agreement with each organisation, no named data can be published in any form.

3.8.9 Summary

This Phase contains the whole point of the Research for the doctorate process, from the creation and use of the idea generation framework, to creating the framework (which is outlined), and the explanation and justification of the Case Study method. In effect, this is the heart of the research, and the results would determine whether the resources, money, and time expended in both the creation and testing of the evaluation framework, as well as the overall learning journey, was to be seen as worthwhile.

3.8.10 Research Phase Four – Reaction to Case Studies

Phase	Design	Deployment	Analysis
Phase Four 2016 To review the attitudes of those taking part in the Case Study	<u>Sample</u> <i>Purposive Sample of all 6 people who were the key L&D practitioners involved in the process</i>	<u>Survey</u> Questionnaire sent to all 6 respondents. All 6 also interviewed by a third party	<u>Content Analysis</u> Content analysis using proprietary software – converted to Excel for reporting in this document

Before beginning this process, due regard was given to all the factors covered in Phases One and Two which had also been created using survey design and collection. For the sake of efficiency, all stages regarding question design, delivery, and analysis were replicated, the

sample size is only six people, and most the choices were straightforward. Having completed the Case Study and analysed the returns, the final stage was to collect the opinions of the L&D stakeholder in each of the research organisations, to check attitudes and opinions regarding the performance of the framework.

Following a discussion with each of them, a decision was made to restrict that research only to the actual L&D client and not to delve deeper within their organisations, choosing to respect their opinion that they would be best able to judge the credibility of the process. Whilst I would have preferred to judge the credibility of the process from the perspective of the customer of the L&D function, I accepted the decision and, in some cases, helped them devise suitable questions for each of their own organisation line managers to deploy as they saw fit.

Therefore the final questionnaires covered any perceived gains, opinion on the process, and any other feedback that may be forthcoming. A short questionnaire was devised, questions were again run through the internal pilot group I had used for the other parts of the process, and the survey was deployed. I did consider interviewing each person, but decided that the risk of having broader discussions, or diffusing into wider points of 'fixing' some of the factors that were affecting their actual ROI results, should be dealt with separately. Therefore, the short survey was seen as quicker, and more likely to garner more frank contributions. The response rate was 100 per cent. Again, blending discourse analysis and content analysis using our software, a series of output charts was created, and are presented elsewhere in this document. However, there were very few questions, as I was looking for a simple series of answers to clear questions so little analysis was required.

3.9 Ethical Considerations

This project builds on the successful completion of a number of client projects, and subsequent professional relationships with suppliers,

clients, and other training providers, conducted by Russell Thackeray as a Director of QED Evaluation Ltd.

Having reviewed my own position with regard to ethical issues, my operational position will continue to be informed by principled relativism or 'situation ethics' (Fletcher 1999). Essentially each situation will be dealt with utilising the appropriate ethical position on a 'case by case' basis as espoused by Goode (1996).

QED Evaluation Ltd, my consulting practice, operates a similar perspective depending on client needs, which range from Small to Medium Enterprises (SME's) to the National Health Service (NHS). In addition, the company operates its core processes against the ethical and contractual frameworks operated by the following bodies:

- Institute of Consulting
- Chartered Institute of Personnel and Development
- British Psychological Society
- Chartered Institute of Marketing
- Chartered Institute of Management
- The Evaluators Institute

There are a number of ethical criteria when constructing a project including, but not limited to:

- Data Protection
- Confidentiality
- Maintaining professional relationships
- Retaining professional objectivity in the face of commercial pressures

There are a range of stakeholders whose needs are also considered during the research processes:

- Delegates
- Other Training Providers

- Operational Managers
- Programme Sponsors
- Budget Controllers

The research for this doctoral programme builds on ethical criteria, which is operated as normal practice in QED, in order to reinforce good practice, and ensure no dilution of commercial confidentiality. Given the high level of ethical and commercial standards currently in place, I propose to replicate these organisational steps and ethical steps:

Area	Steps Taken or Proposed
<i>Awareness of the project</i>	<ul style="list-style-type: none"> • Every client has been contacted and has verbally agreed that their data can be used in relation to the doctoral programme. However, absolute confidentiality has been agreed for them within the writing of the report and any other written material • Assurances regarding commercial sensitivity and confidentiality have been made. • No local or national government organisations will be part of the doctoral research. • No external body is required to grant ethical approval for this doctoral project
<i>Data Protection</i>	<ul style="list-style-type: none"> • All data is saved in QED storage for a period of ten years following the completion of the project. • All client reports are saved for a period of ten years following the completion of a project. • All tape recordings of meetings and meeting notes are transcribed within a period of six months following the interview and are then destroyed within six months of the transcription date • Any completed surveys or emails or any other form of correspondence from contributors will be destroyed totally after a period of one year

	<p>following the completion of transcription</p> <ul style="list-style-type: none"> Any data or outputs will be owned by Russell Thackeray explicitly for the purposes of completing the doctoral research Any subsequent marketing activity will be agreed individually with each client
<i>Confidentiality</i>	<ul style="list-style-type: none"> Delegates are never identified by name – only by a reference number or a function (where the function has more than 10 members) Companies are not identified by name unless they have given us explicit permission to do so. In most cases, QED have a Non Disclosure Agreement in place to protect all commercial and personal sensitivities and confidences QED do not publicise the results of clients in their marketing material QED create benchmarked or normed ROI figures for clients – these are only shared with companies where agreement has been reached, and are not marketed for general consumption One of the issues is whether a secondary reader or analyst of the data is as trusted as myself. Alderson (1998), suggests that this issue must also be considered both in the analysis, writing up, and publication phases. Any potential issues of confidentiality from hearing information which creates a problem for the researcher will be dealt with through our supervision process, and building on the ideas of Erikson (1967), who suggests that informed consent be a feature of briefing for any delegate or any researcher, and that appropriate ground rules are put in place with regard to what may or may not be the subject of confidence

	<ul style="list-style-type: none"> • QED has those ground rules in place – in this case, we would not be asking about any areas which impact into personal areas or actions other than the outputs of training courses. Anything, which the interviewee asks to remain private, will not be captured. • Interviewees are always offered a transcript of the meeting to reassure them that any confidential material has not been captured.
<i>Maintaining Professional Relationships</i>	<ul style="list-style-type: none"> • QED operate a tiered process which means that an Account Manager works with the delivery team to ensure that all meetings, feedback, and reviews have more than one person present at all meetings • QED employment and commercial contracts establish standards of behaviour with clients
<i>Interpretation of Findings</i>	<ul style="list-style-type: none"> • All findings are analysed by a different team from that which collected the research to remove any emotional 'pollution' from the interviewer in the process – this is an organisational standard, however the work was carried out by me for the purposes on the research • All findings are fed back to the client, as they have been collected irrespective of feelings/ego of the receiving client • A second part of the process is to manage the impact of the results and findings with the recipient (this utilises a consulting skill set) – the different parts of the process are handled in one meeting by different people to mark the difference in the content and the stage of the process
<i>Appropriateness</i>	<ul style="list-style-type: none"> • All materials are constructed using plain English. They can be translated into any language, at the

	<p>request of a participant.</p> <ul style="list-style-type: none"> • Any meeting can be conducted in the language of the participant's choice • Where visual or auditory impairment is an issue, specially trained people can help, or material can be constructed to aid communication. Any feedback is also handled using the person's preferred method of communication (We work for a number of disability charities and conduct surveys for them against their own standards) • No materials are intended to create any form of offence deliberately or otherwise. A complaints and appeals process is agreed with clients in advance of any project. All questions and feedback are checked with this in mind. • No research will be carried out with any person under the age of 18 years old • No access to personal or confidential records is required for this doctoral assignment • No specialist psychometric instruments will be used in this project
<i>Reporting</i>	<ul style="list-style-type: none"> • Access to a written and verbal report has been agreed with those people who want to take part in the project • Access to the written report is available to anyone who was contacted as part of the research, irrespective of whether they choose to contribute or not

Within the Research Project Report there are references to the use of third parties. It is important to be specific about what they are and what they did and why and to justify this on (particularly) ethical grounds. Gray (2009) advocates four ethical principles that need to be considered

and reviewed to ensure that the research does not create a breach of ethics. They are:

- Avoid harm to participants
- Ensure informed consent to participants
- Respect the privacy of participants
- Avoid the use of deception

Whilst the first three are covered elsewhere, it is important to ensure that the fourth point is not directly or indirectly compromised within the research. The use of third parties is understood in the medical and scientific community, but some attitudes and approaches need to be explained in the context of social sciences research.

Data Entry Checking

During the data input stage, a colleague within my professional practice (acting as a critical friend) checked the data input to ensure accuracy. They were fully aware that this was an academic project and that I wanted to ensure that it should operate with full robustness. This was a sensible precaution as my ability as an evaluator rarely extends to the operational handling of data. Therefore I inputted the data and it was simply checked by the third party. That person was fully aware that their role was to be none other than a check for accuracy and they had no access or opportunity to conduct any analysis or interpretation. The person also operated a check so no typographical errors had appeared in the coding frameworks as well. In effect, their contribution was no more significant than a basic secretarial service.

Panel of Experts – Use of employees

The work of DiPiazza and Eccles (1992) who suggest that the 'Delphi Method' as well as the 'Consensus Method' be used when planning a

forecast. This process uses a number of experts to review and forecast an outcome with the benefit of expert knowledge. In addition to this both De Bono (1984) and Kirton (2003) recommend the use of wide reading of Delphic influences outside of the specific confines of a research subject to stimulate fresh thinking and enlightenment.

It struck me as an appropriate use of experts within the field to ask them to operate as a 'panel of Delphic experts' to challenge my thinking and assumptions around the basis of the forecasting approach. The ethical position of the time spent together is shown in greater detail during the document but their role was simply to pose challenging questions in order to add value. As experts, they were also asked to generate an estimate of a best practice benchmark so that I could use this to determine a level of accuracy needed by the research. Their expertise was accepted and used as the basis for the initial back testing approach shown later rather than relying on my own personal opinion.

A number of the panel were employees of the evaluation practice I work for. Whilst this may be seen as unusual, each of us are members of professional bodies and adhere to those ethical standards. They are each senior and respected practitioners, used to presenting difficult and challenging reports and findings to a range of extremely senior and challenging clients and funding bodies. None of them has any issue with presenting their views or findings in an open forum. In order to ensure that this was the case, a facilitator was used to ensure that each person was able to contribute fully as necessary.

In the same way that it is possible for a manager to coach or counsel a member of staff by utilising the ethical considerations of the appropriate professional qualifications and professional body, it is possible for qualified, highly intelligent, published consultants to be able to work together in a non hierarchical manner.

This type of working was enabled by virtue of the company operating as a small consultancy practice. Kaplan & Norton (1996a) and Gummesson

(2000) comment that within smaller companies a culture exists that is different to that found in large institutions or corporates, and one which promotes the open discussion of work without fear or favour by senior highly respected and remunerated consultant such as those that work for QED. Regular reporting in the trade and media refers to the cultures in start-ups and small entrepreneurial companies and this should be borne in mind when thinking about employees.

External evaluator

Within the literature review contained in Chapter 2.9.2 is a discussion of the areas of the literature concerned with the criteria needed to validate the effectiveness of a forecast.

Walton (1993), Armstrong (2001) and Seigel (2013) all stress that the goal of a forecast is to be able to match the results generated from an existing summative process. This process of 'comparative' or 'rival' forecasting is a 'highly desired' one. This process is often used in the medical and scientific forecasting arena where, for example, a number of labs will carry out a project using an existing approach and this is then compared with the new approach that has to better or improve on that result. It is commonplace for the other parties to receive a commercial rate of remuneration for their work to ensure they operate in a commercially robust way having to observe their industry and ethical codes of practice.

This is the process deployed in this research project. An independent professional evaluator was engaged to generate a summative result using the most robust and respected Phillips method. The evaluator was paid a commercial rate to ensure that they operated completely independently of this research and against the standards and requirement of the evaluation professional body. They had no sight of the research in this document before or during the project and no vested interest in the project other than to provide a legitimate and robust result that would be

seen by the organisations involved in the Case Study. Not to have paid a commercial rate for the work would have led to an accusation that a favorable result or a less rigorous process would be operated by them. This is different perhaps than the social sciences where the implication of payment produces a skewed result. In this type of research, working 'as a favour' would have created a greater ethical problem.

Using this approach meets the test in forecasting research requirements, case study methods as well as avoiding any breach of the Grey (2009) criteria.

3.9.1 *Summary & Conclusions*

The initial research proposal was developed and amended to ensure its relevance, and to reflect the learning developed from both the academic literature, as well as the unfolding of practice.

The research was designed to answer a number of questions, identified earlier in this chapter, and the results are contained in Chapter 4.

Using an opinion survey, a more focused diagnostic survey, and a seven organisation case study has created logistical problems in terms of delivering the research project within the realities of day-to-day work and life pressures, and has unduly extended its life and duration. However, this has helped with the emergence of both phenomena and meaning and has created a new framework as well as tool-kits that could add value to practitioners in the field of L&D.

The inevitable compromises, within each of the sections and stages, create the opportunity for more research, and a different analysis of the data but, given the parties involved and knowledge and resources at the point of the research being conducted, it could be said that the research should be robust enough to allow for a cautious proposal of a new framework.

The next chapter will compare findings from the research phases, as well as the literature study, to investigate opportunities for future research, as well as guidelines and recommendations for those of us engaged in the practice of evaluation.

Chapter 4: Research Project Findings

4.1 Introduction and Context

This chapter shows the findings from each Phase of the research by illustrating the key outputs from each phase with some commentary, clarification and extra detail where it serves to add value.

The initial thinking for this research was to create a paradigm shift in the practitioner base by creating a new framework. Initially, the idea of the research was to carry out a large-scale opinion survey to assess attitudes and to see whether a new framework would be conceptually possible.

One of the issues with the research was the length of elapsed time between phases. This was due to the learning process as well as external factors and this created challenges for the efficiency of the overall process; however, it did create opportunities to discover learning and meaning, and allowed the research to flex and adapt to the creation of new processes and ideas within evaluation practice. As ideas and framework creation emerged from the initial research, practice and literature review, the primary practical focus of the research became the Case Study outcomes in Phase Three.

The challenge with Phase Three then became the somewhat binary question of whether the framework performed within tolerable degrees of accuracy to the benchmark standard, at a substantially reduced cost of process. Whilst this may be less satisfying in terms of academic research, the simple findings represent almost a revolution in the way practice can think about evaluation. The final Phase was an adjunct to the Case Study and became simply the collection and capture of attitudes to the Case Study outcomes.

The research phase covers four phases that combine to form the whole:

Phase	Overview
<i>Phase 1</i>	An initial survey of attitudes within a purposive sample of

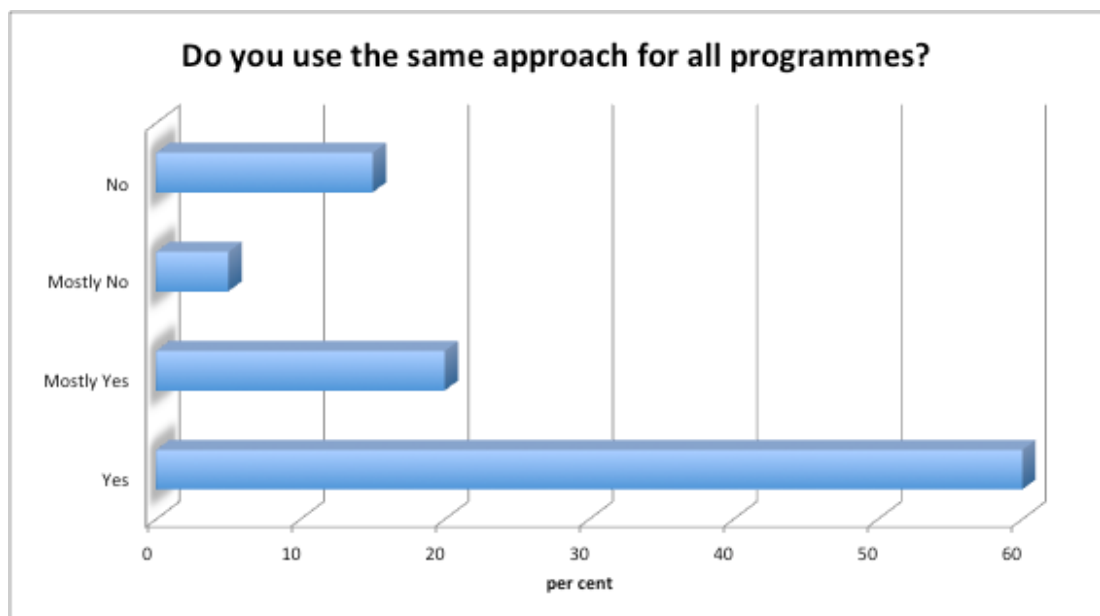
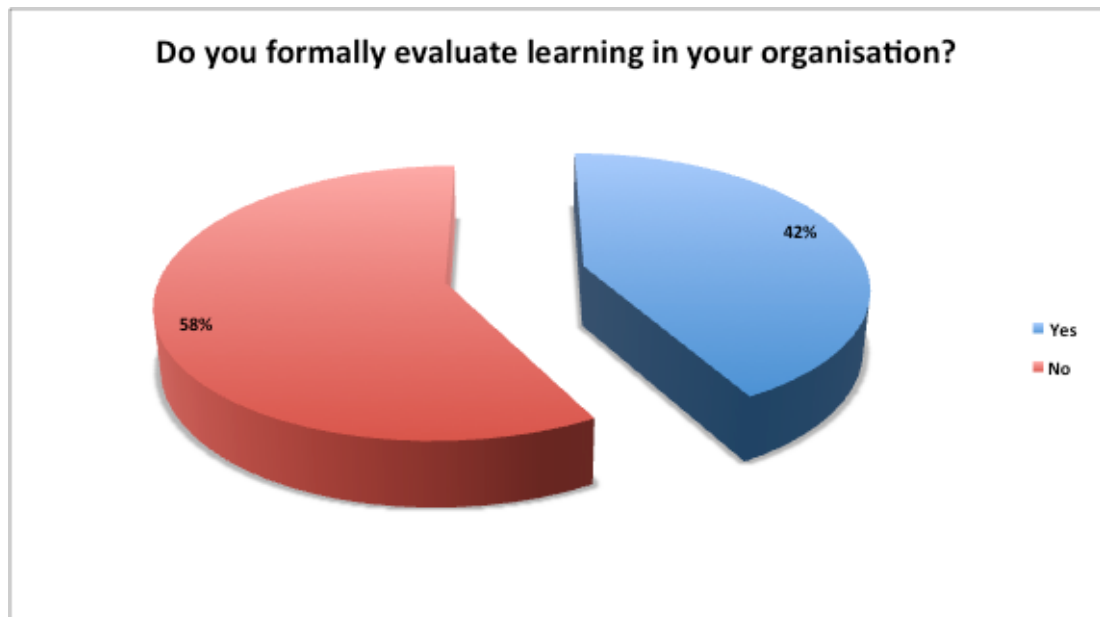
	relevant L&D professionals
<i>Phase 2</i>	An investigation into the practice and operation of a smaller group of evaluation practitioners
<i>Phase 3</i>	A Case Study review of the new process split into three parts <ul style="list-style-type: none"> • a test of the process on a retrospective sample • comparison of the process outcomes on a range of courses relative to the 'benchmark/gold standard' method • comparison of costs of both profiles to determine whether any cost savings justified any major variation in the results between the processes above
<i>Phase 4</i>	A survey of the views and opinions of those L&D professionals that took part in the Case Study research

In each case, the findings from each Phase are presented or summarised to highlight key findings with some clarification of process, or to point out items of interest or curiosity. Each section of findings is illustrated by simple diagrams, these have been translated from our software into Excel charts for the sake of consistency and ease of reading. Every person was required to complete every question.

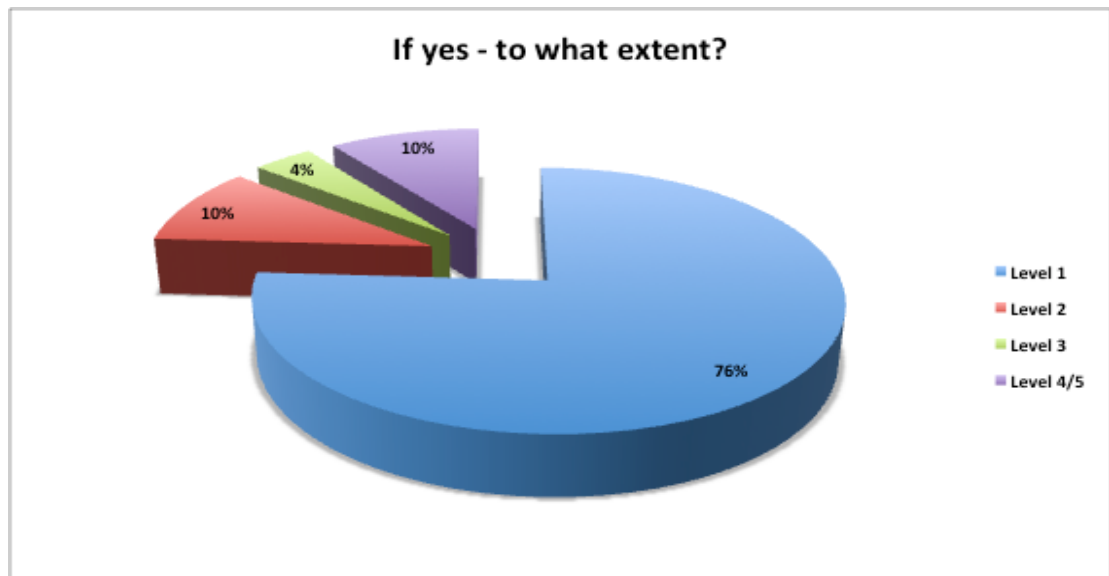
4.2 Phase One — Results

This stage consisted of a short opinion survey asked of a large sample, as outlined in the Research Methods' section. The overall list of questions is shown in Appendix Two (a).

All of the charts illustrating the responses are shown here.

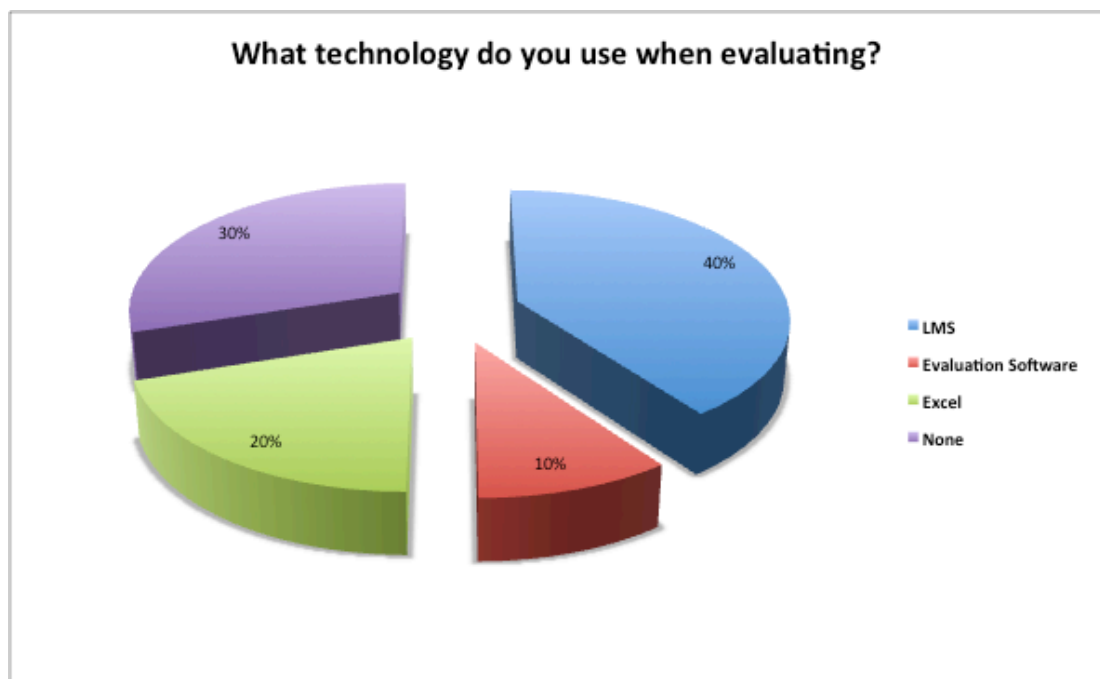


In terms of percentages, this is different from the findings of the most recent CIPD (2015) survey, where they report that there are 14 per cent of organisations not evaluating any form of development. This could represent data-skewing and sample sizes in the respective research. Also the initial starting point for the research is different. The CIPD is a professional body and one wonders the extent to which people want to 'admit' this to such a body. Our research is focused particularly around evaluation rather than a broader look at L&D issues.

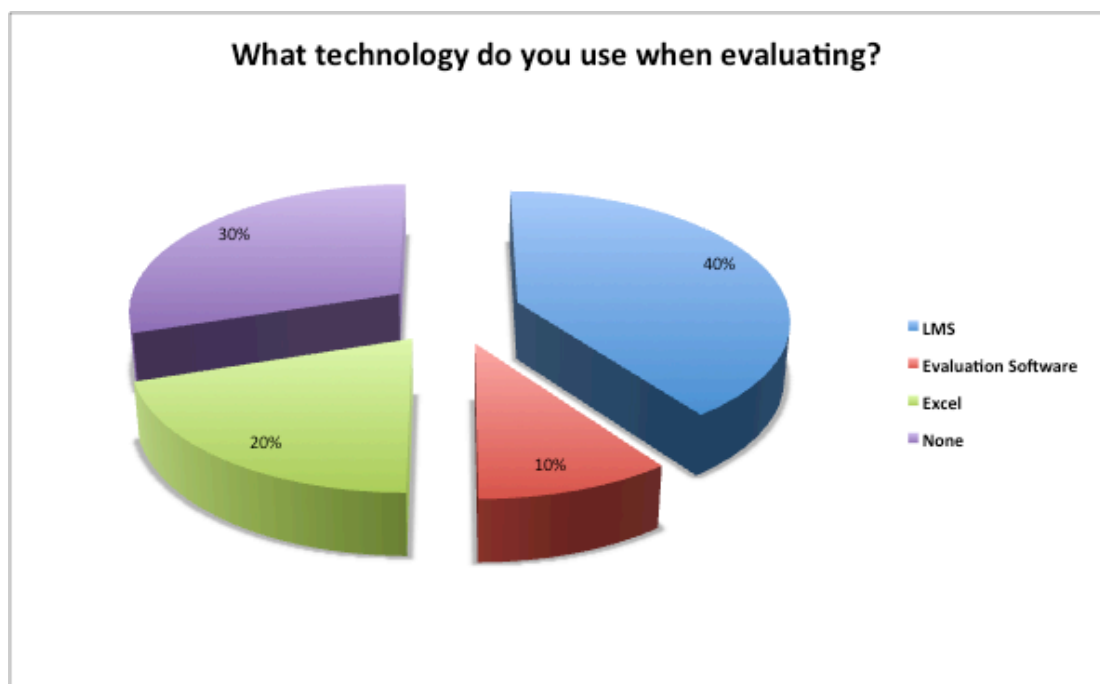


All of the people in the survey appear to have understood the reference to 'Levels', which is a fundamental part of the Kirkpatrick concept. Whether they understood what the detail of what each level actually meant was not covered here.

An observation from one of the participants was relevant here: "Given we only operate to Level 1, there can only be one method applied"; however, the lack of difference in approach begins to highlight the problem of programme complexity being evaluated by a 'once size fits all' evaluation solution.

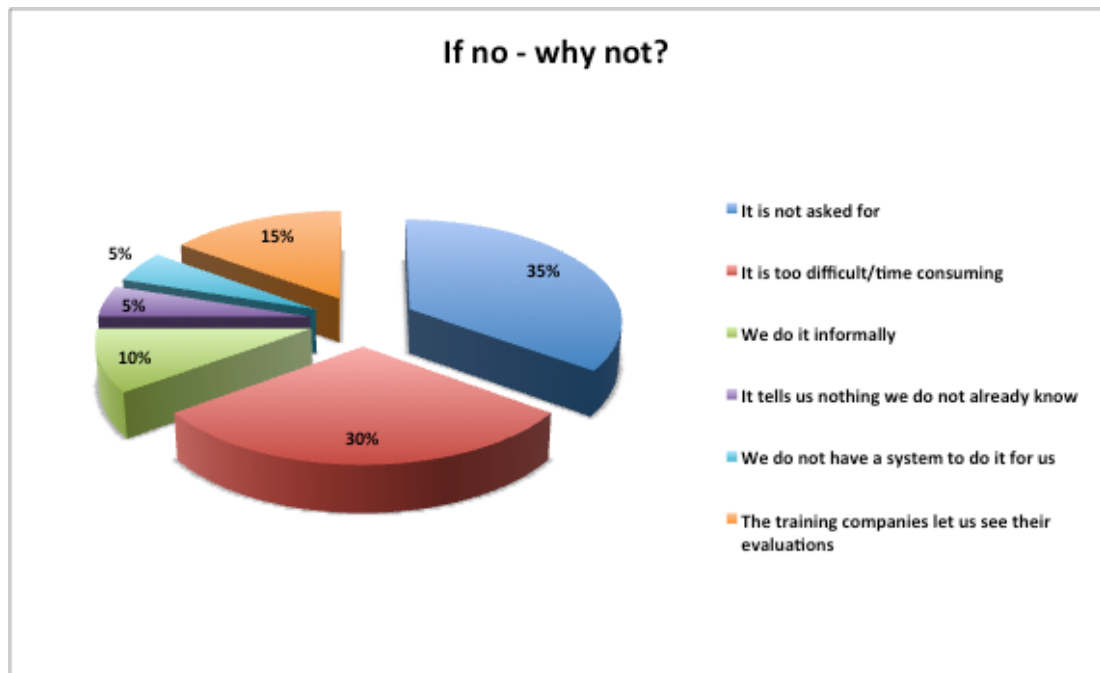


The data shows here a heavy reliance on the use of a LMS either to determine the process or the questions within the process. As the majority of LMSs use the Kirkpatrick method, the LMS is effectively choosing what evaluation is conducted by default. Analysis of LMS websites in the literature shows the use of evaluation data as a means to drive down the cost of evaluation rather than to determine value. This may be a serious opportunity cost for those people who adopt this approach.



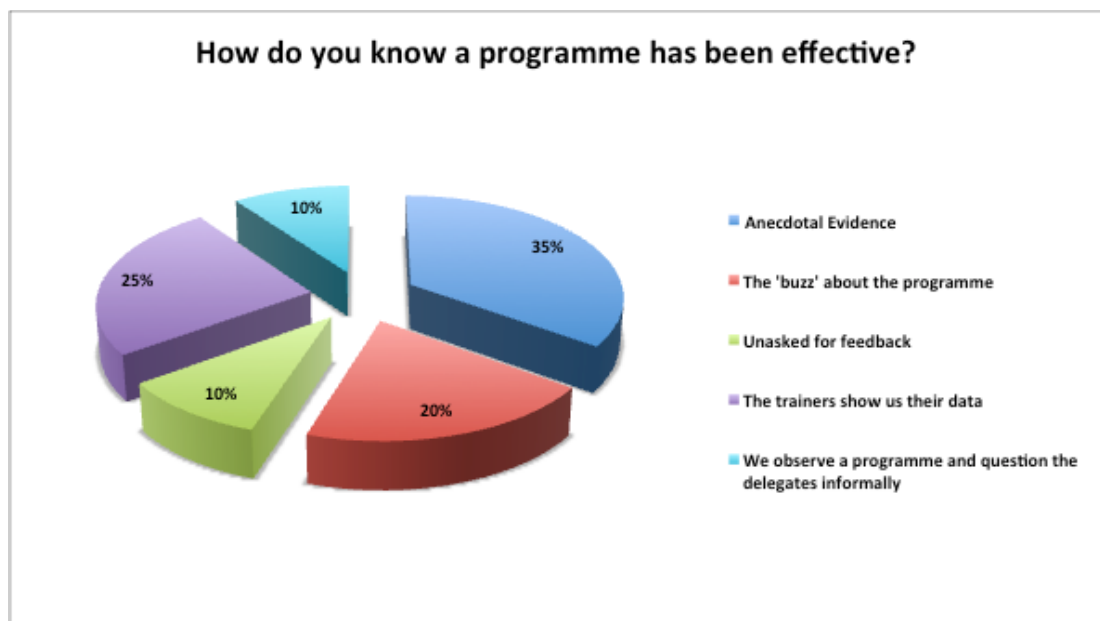
The commentary about having specific evaluation software reflects the process of adding evaluation software to an LMS, or where the LMS can export data to a specialist evaluation analysis firm who can then provide appropriate data and reports.

This research was not concerned with LMS penetration, rather simply attempting to understand how much the LMS drove the process. However, one person commented that they did not use the LMS, preferring to use specialist software; otherwise they "got what we are given, rather than what we need".

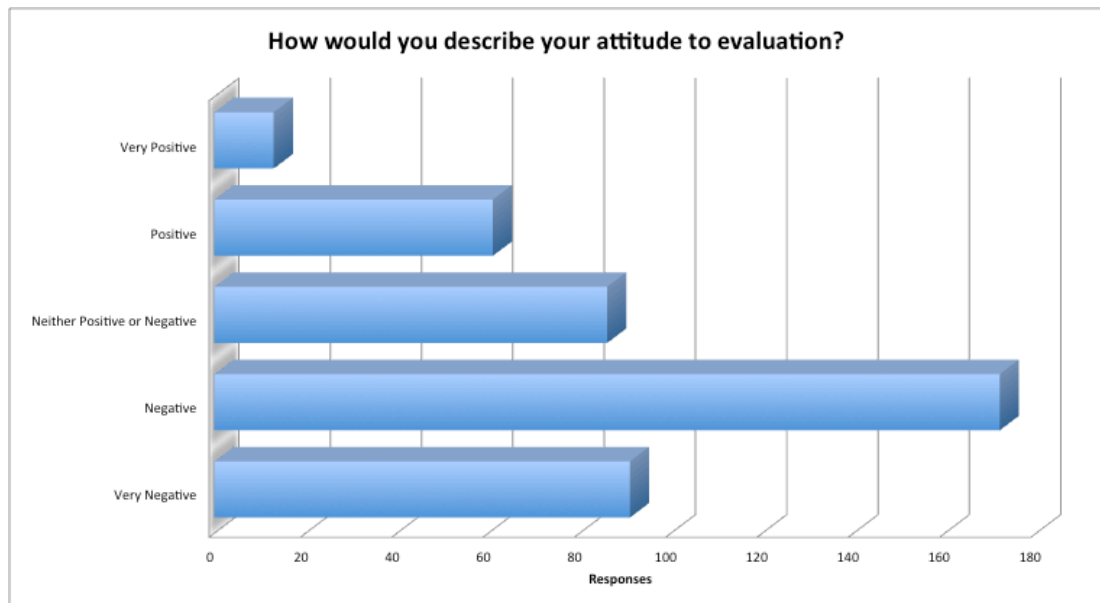


In effect, these appear to be the most important barriers to evaluation and are mentioned in literature, as well as echoing my own experience with L&D functions wrestling with evaluation.

Whilst a small group of respondents are reporting that, although they answered 'no' to formal evaluation, they are doing some form of post - course review – even if not formalised evaluation, sometimes, using informal methods, such as anecdote collection, or asking the training provider to collect the data.



It is surprising to observe how informal the process of evaluation actually is, given the level of spend in this function. The fact that most evaluation is at Level 1 and relies on informal commentary, gives concern as to the value that the training budget actually brings and how valuable it is within an organisation. It also may weaken the evidence of the value of L&D, given the actual nature of evaluation practice within this sample.



Given some of the answers shown elsewhere, as well as barriers identified, it is understandable that most of the views are not positive. This is surprising, given the nature of the group, where everyone had been interested to join in the research in the first place. Perhaps this represents a view that evaluation itself can be important whilst people are still not being positive about the process and methods used in carrying it out.

Other Comments

A range of comments was collected, in addition to the short number of questions – they tended to reflect the same themes:

- We would like to but cannot
- We produce plenty of data, but that does not stop our budget being cut
- A necessary evil
- I just do not have time

- Our budget does not seem to be affected by our results
- I know we should - it just does not interest me
- There is no simple evaluation tool-kit that allows us to do it quickly
- Producing ROI figures is very expensive
- The organisation does not trust our own evaluations
- No-one takes our reports at face value – so why bother?
- Finance do not believe that intangibles can be measured at all
- The CEO believes in training so we have no problems with resource or budget

4.2.1 Conclusion

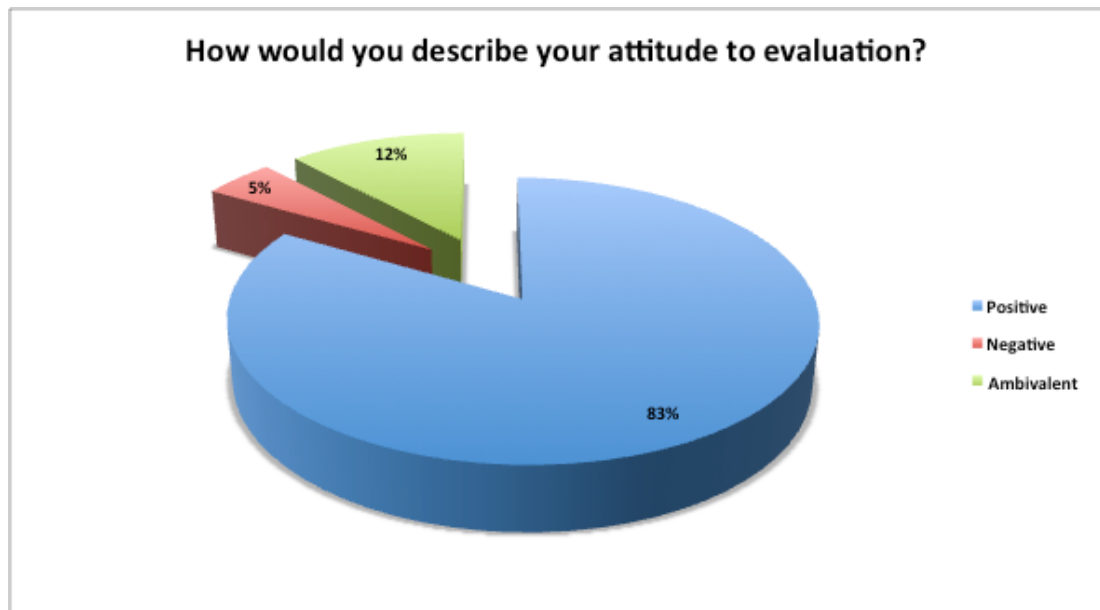
What comes through as the prevailing attitude in this phase was the nature of the difficulty of evaluation (relative to the outputs it delivers). Whilst many L&D practitioners state they are not being asked for the responses, there is a number of trends that may change this result including:

- Potential market conditions for training investment post-Brexit
- Attention from external regulatory bodies now including training impact as part of their audit regime
- Attention to the subject from Investors in People etc

However, the attitude of difficulty appears to be relieved by the use of the LMS to create a simple solution; even if this solution could be part of a larger problem as time progresses.

4.3 Phase Two — Specialist Practitioner Research — Results

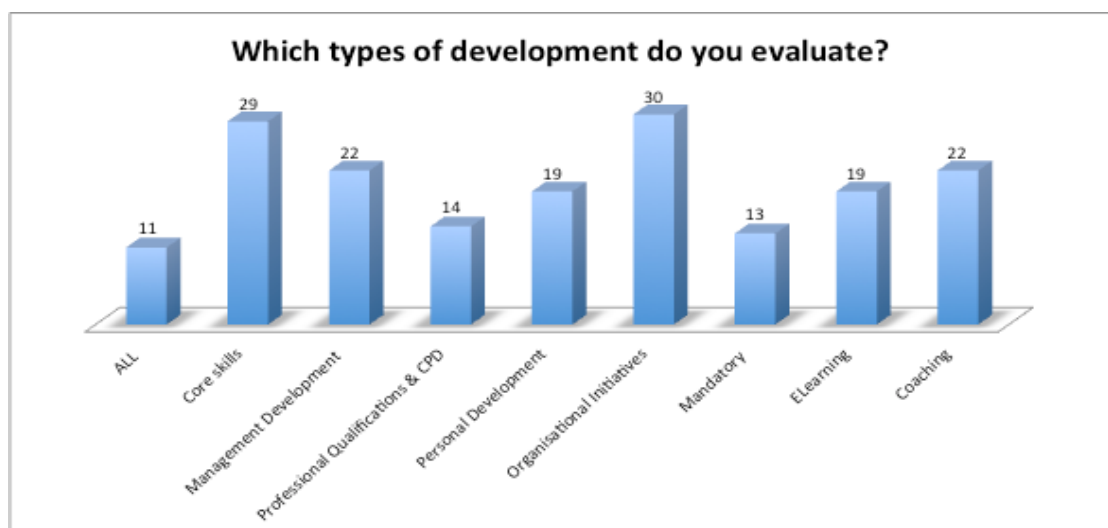
The charts shown below are drawn from the 34 questions that were posed to the research group. It is worth pointing out that members of this group were all selected and were willing and motivated participants in the research. Each of them had expertise in the subject with some being highly qualified in the subject. The overall list of questions is shown in Appendix 4.



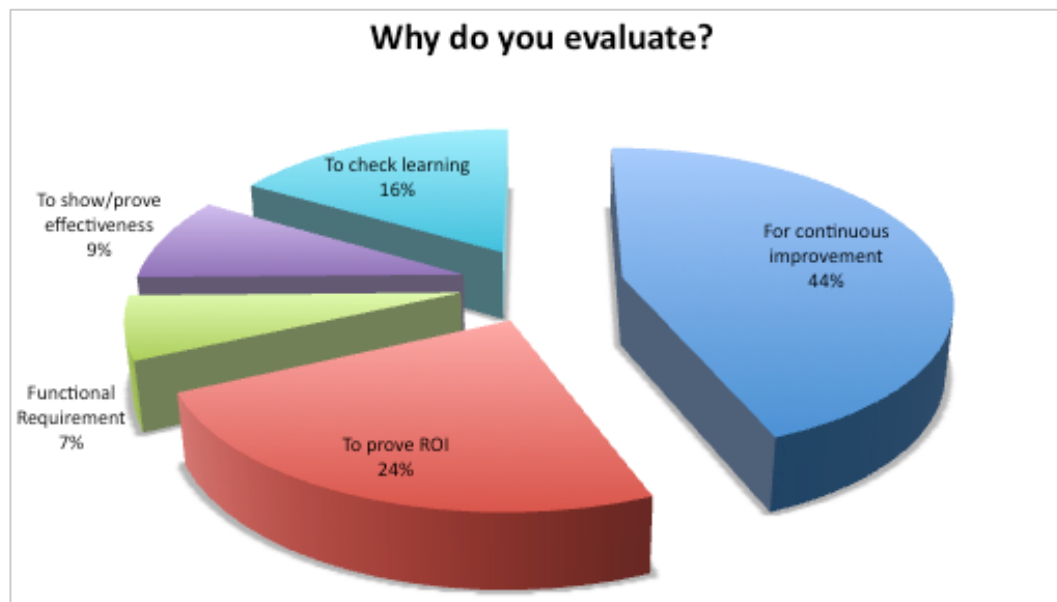
Relevant comments:

- It's the most important thing we have to do
- Something I prefer to leave to others
- A vital part of the L&D mix
- It is something I am expected to know how to do
- It's the Holy Grail
- Anyone spending budget should be expected to account for it
- Everyone should do it

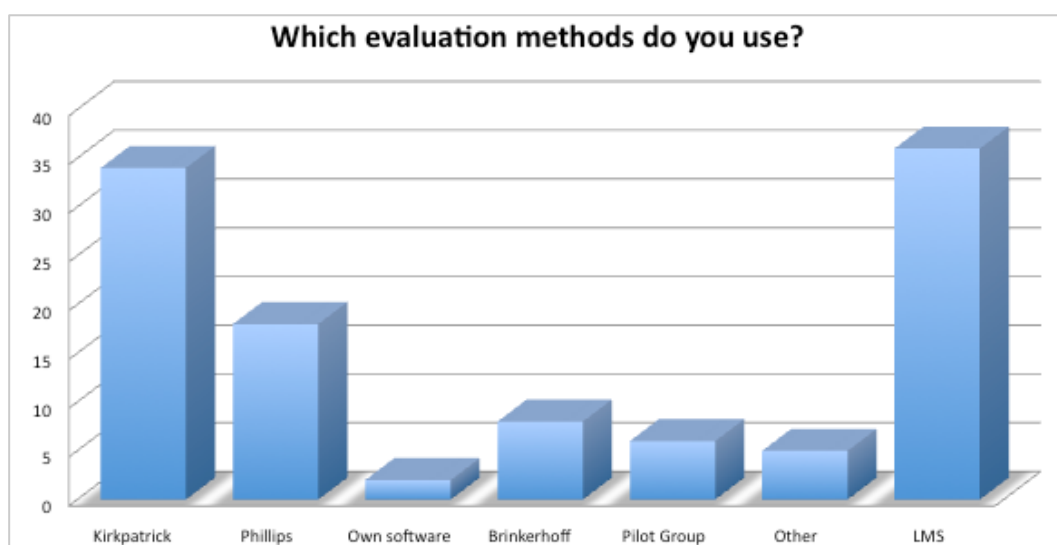
The level of positivity in this group of practitioners that actually conduct evaluation is in stark contrast to the previous Phase and may reflect the fact that skilled practitioners have more confidence in the process and outcomes of standard evaluation processes.



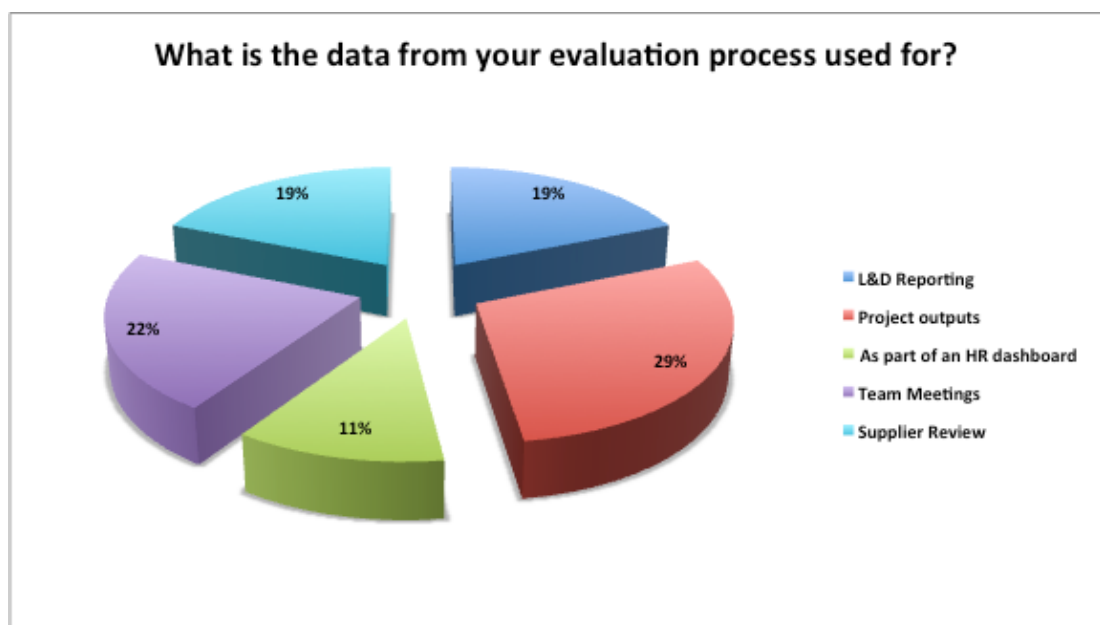
Those that answered 'All' (11) had their answers added to those that only selected a particular category. The level of mandatory training evaluation is low, often reflecting the repetitive nature of the development. In addition, as the development tends to be e-Learning based, the resultant evaluation tends to be focused on knowledge acquisition, rather than impact.



In the comments section, 15 recipients stated they evaluated to assess the effectiveness of a supplier and 10 added in the need to check that the development was meeting the needs of the learners. It could be argued that this is the practical use of evaluation, as part of a wider decision theory approach, rather than seeing evaluation as an end in itself.

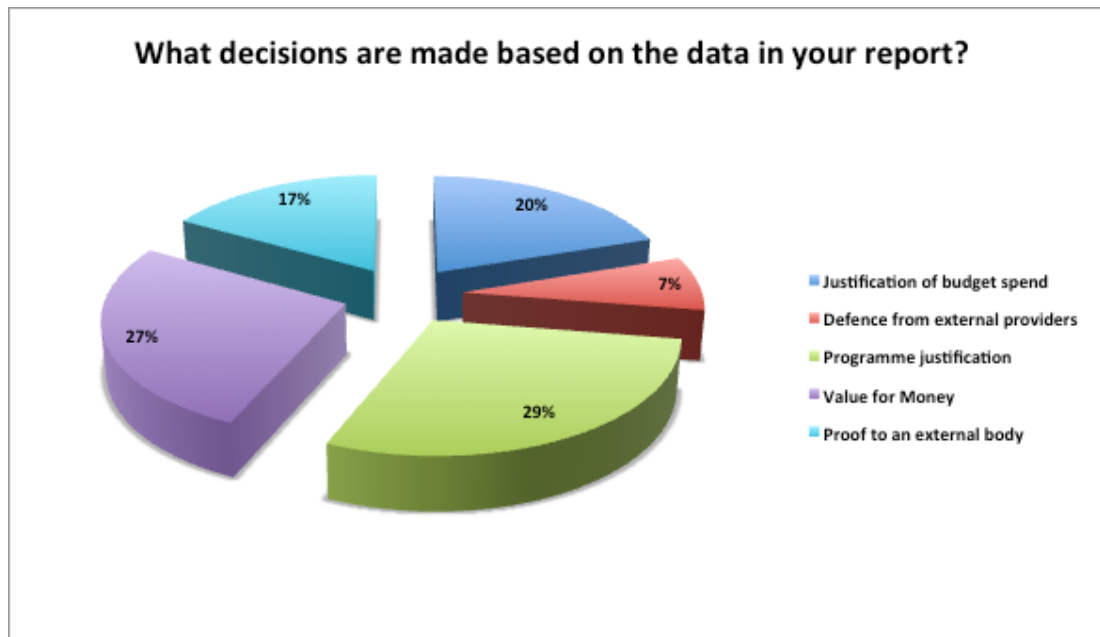


Some of the participants use a combination of different methods, dependent on what they are attempting to find out. This shows a degree of sophistication, sometimes not desired (or possible), when only using the method available within a LMS. However, because the Kirkpatrick method tends to be the framework of choice for LMS suppliers, adding the majority of the LMS returns to Kirkpatrick reinforces it as the prevailing paradigm. The rise of NPS is fascinating, and it would be interesting to investigate the levels of understanding of this measure, and how it relates to training, given the findings in the literature review that learning and satisfaction are not linked.

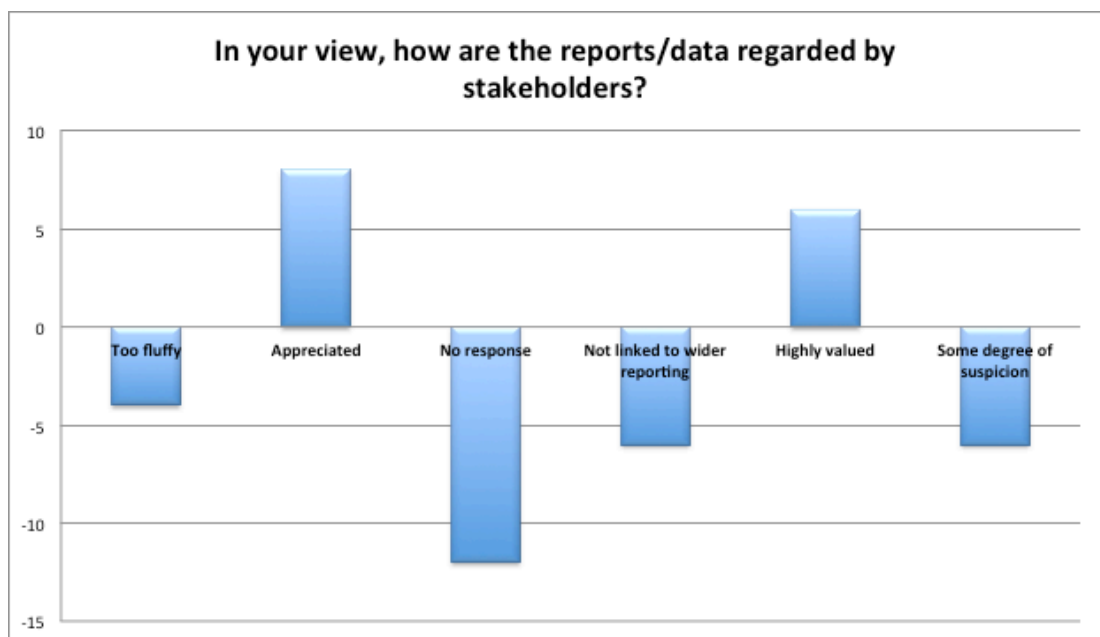


Not all of the respondents produced reports. For those that did, it is interesting to see that where L&D results are used as part of a wider justification of an HR dashboard, where there may be a link to decision theory, about the value and justification of the wider HR department.

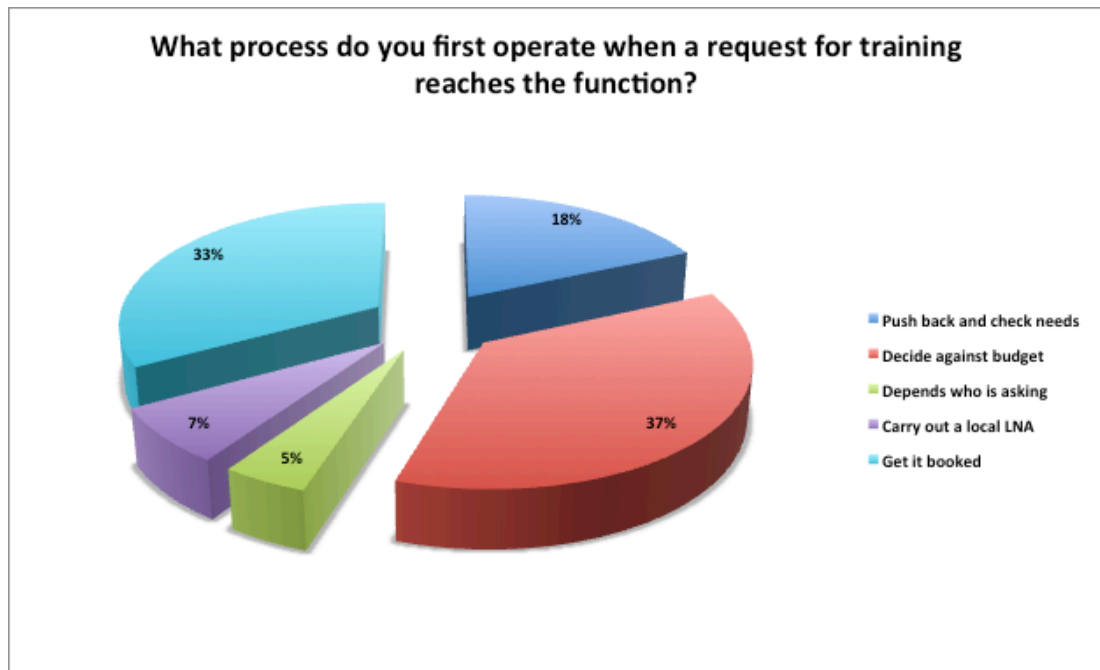
Encouragingly, where reports are produced, they are often being used as part of operational decisions to inform progress and practice, rather than simply as a set of inert data.



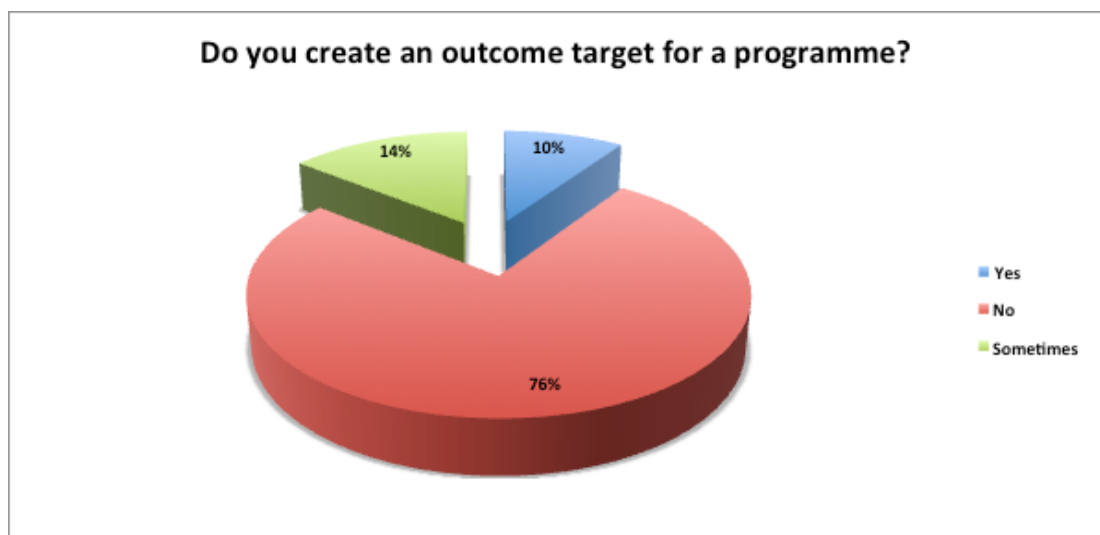
This is clear evidence of Decision Theory in action. The use of data as a means of defence from external competitors is the most encouraging use of this approach, turning data into usable criteria for informed decision making and taking.



Negative responses have been portrayed with a minus score and positive commentary with a plus score in this chart. This is clearly an area for improvement for L&D and may represent a problem in reporting or the fact that the data reported is more suited to the internal needs of the function.

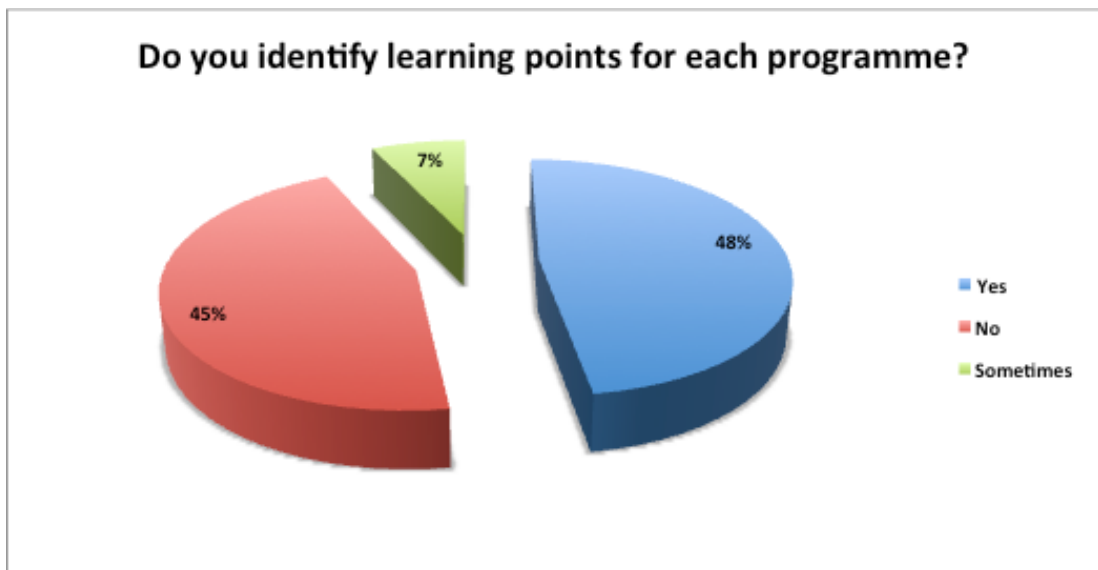


There is a high degree of contextual information that drives this decision and most participants do not have a standard approach, although they do have regular pattern of operation. This idea of 'commissioning' is extremely important as the CIPD L&D trends (2015) reported that nearly 50 per cent of needs are delivered by bespoke training design. A failure at this stage can have compound negative results at each stage of the subsequent process.

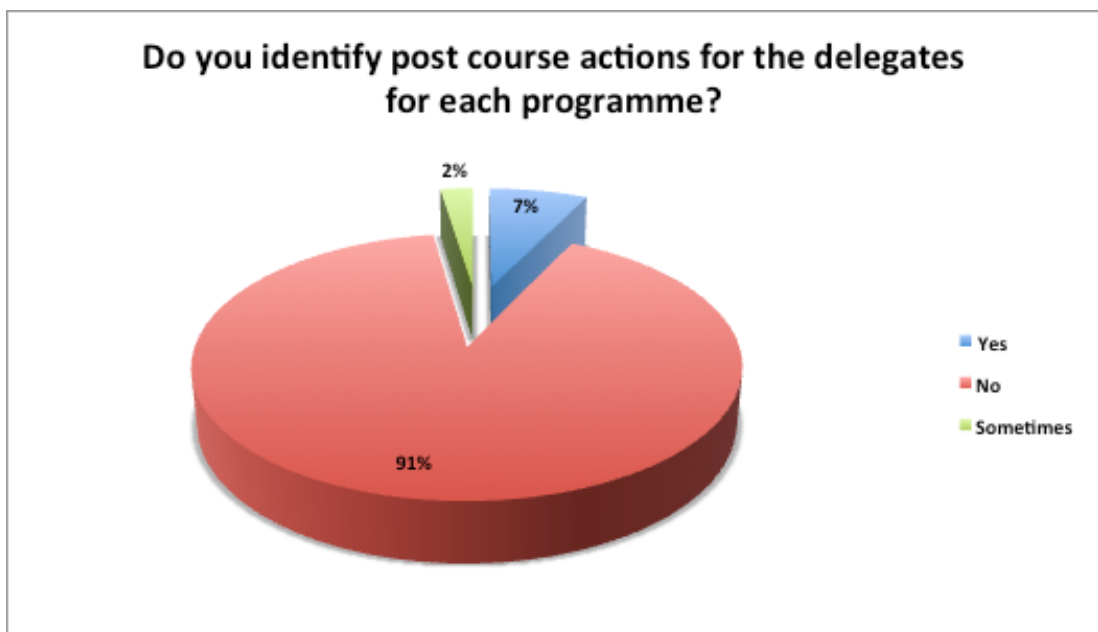


The concept of Return on Expectation seems to be let down at this point. This is ironic, given the prevalence of the Kirkpatrick framework, and that ROE is at its heart. However, it is recognised that L&D may not equate a target with an expectation at this stage. Most report the use of Learning

Objectives as a means of generating clarity around expectations and perhaps that is a fairer way to assess the move to ROE.

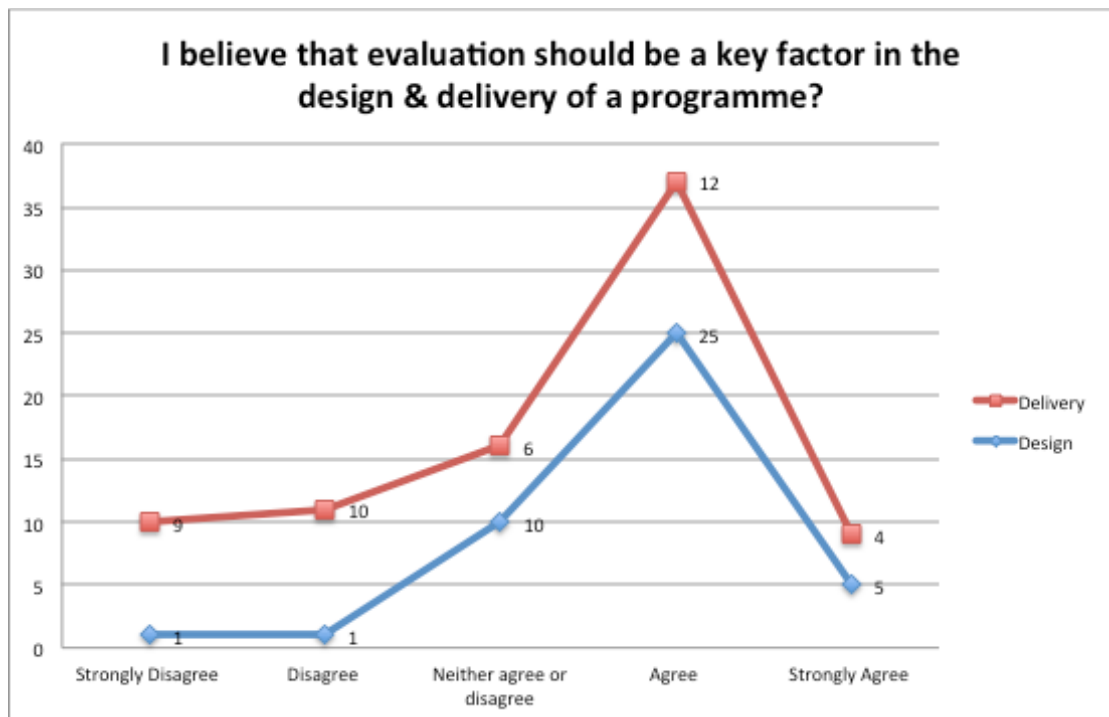


It is within the design process of many organisations that courses have learning objectives rather than learning points. The analysis recognises the difference, even though many of the respondents do not. The majority of the comments referred to learning points as if they were objectives.

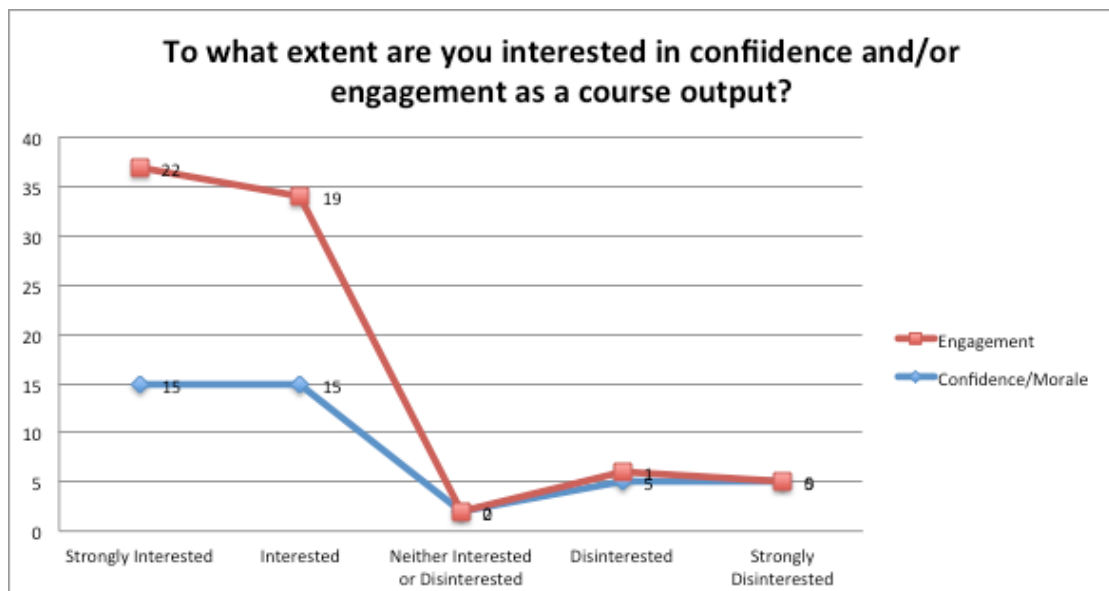


Both of the areas above highlight the level of analytical information not being generated at each stage of the L&D process. This is a problem for training delivery, learning transfer and impact evaluation. One of the challenges here is the extent to which post-course actions are left to 'chance'. As many delegates anecdotally attend courses in the hope of

learning 'just one thing', then much of the learning could be 'scrap learning' or not recognised as important by the trainer or delegate.

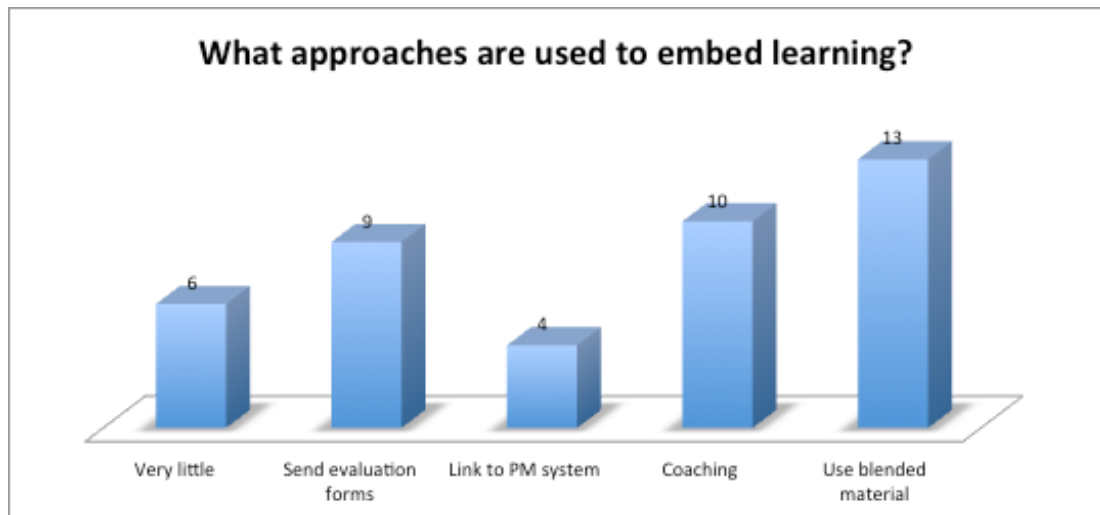


The red line represents the actual responses to the question applied to 'Delivery' and the blue to 'Design'. Both lines should be perfectly aligned.

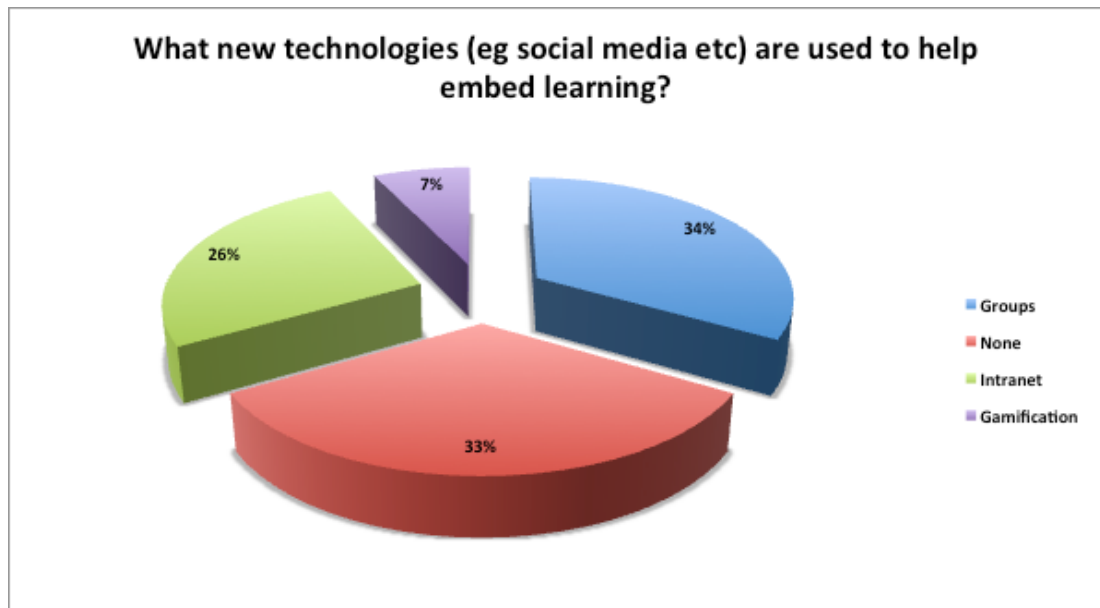


The lack of differentiation here is a challenge, in terms of learning transfer design. Both factors are important for the learner in terms of transfer and not to be clear of the relevance and importance is a point of

concern and opportunity for the function. However, one of the issues here is the extent to which the term 'Engagement' is now linked to the concept based on the Gallup criteria, or the use of normal English without full knowledge of what Engagement is. Many organisations also measure Engagement in staff surveys using their own definitions, or their own criteria, and this may 'pollute' this response.



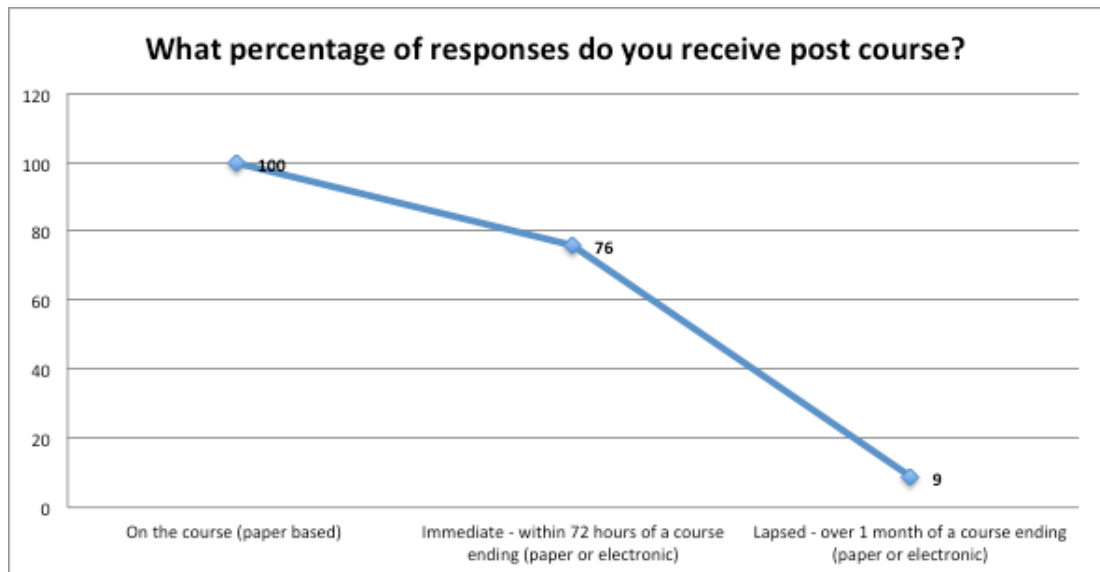
Within our own data (which is outside the scope of this research), we have extensive evidence (using control groups), of the difference in impact for programmes that do, and those which do not, contain coaching as a core element. Not to be using this core L&D approach is a major concern as well as an opportunity for the function. However, the cost of coaching, if used as part of the process, can be a negative factor here. Interestingly, the rise of blended learning, often deployed to reduce 'cost per head' as a training KPI, is an opportunity for better transfer if used more imaginatively.



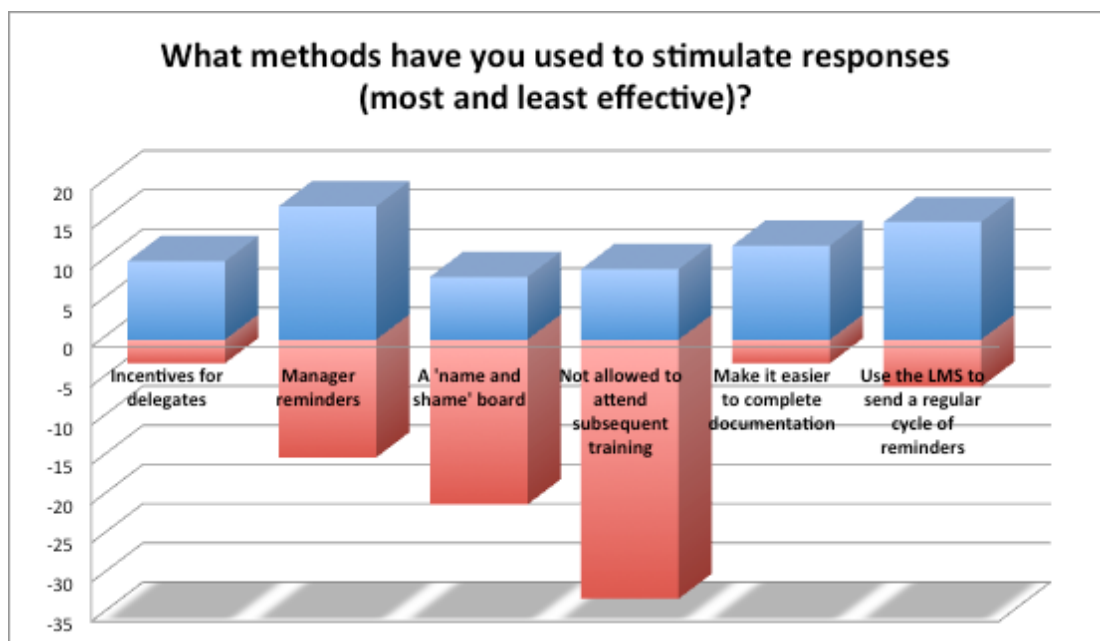
Given the nature of L&D is to be stimulated by new ideas and be open to learning advances, this is concerning. As identified in the literature, there are advances in this area, but few are being used or considered.



Part of learning transfer is to skill managers in these areas. Much work appears still to be done in this respect. Commentary around the complexity of management development and leadership needs takes focus away from this type of management intervention. Little training exists to skill managers in extracting maximum benefit from training courses, although some report they insist that managers sit down with delegates before and after a course to discuss learning and goals.

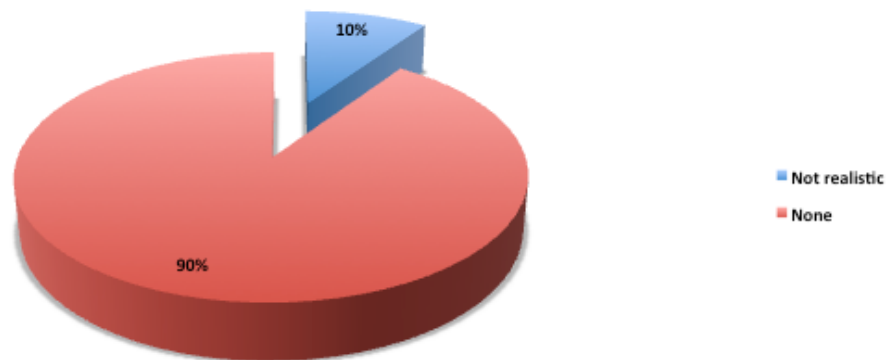


It is of no surprise to see these results reflect other findings in wider research as well as the CIPD (2015). It appears that the move to involve learners in the evaluation process is a major opportunity for the function, as the L&D responsibility to collect data from delegates post-course is a complex and ineffective process. This figure is especially concerning given the penetration of LMS systems and their function to drive efficient post-course response processes. As one person pointed out, improving training response is a 'cultural issue that is affected by survey fatigue'.



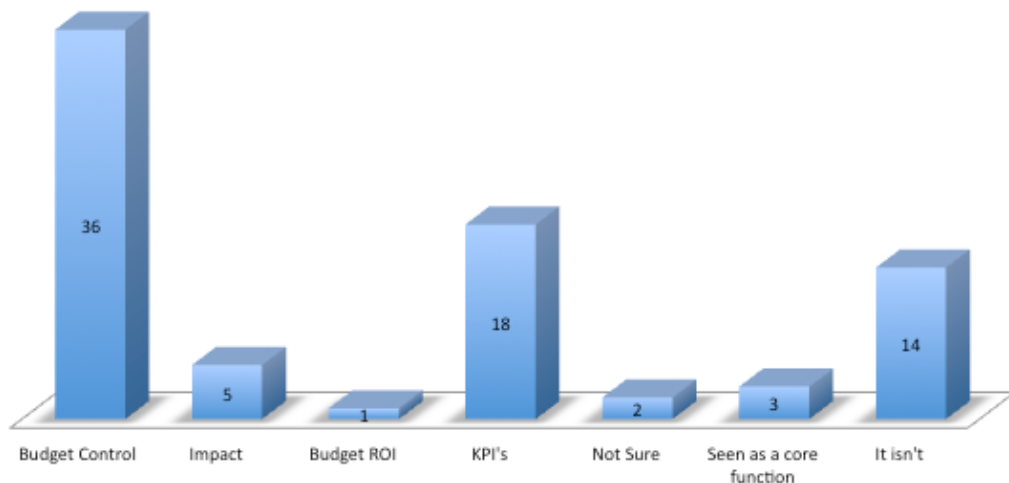
Again, industry wide statistics are repeated here. The perception of little innovation and change continues to be borne out.

What methods have you used to foster a culture where the delegates voluntarily evaluate their own learning?



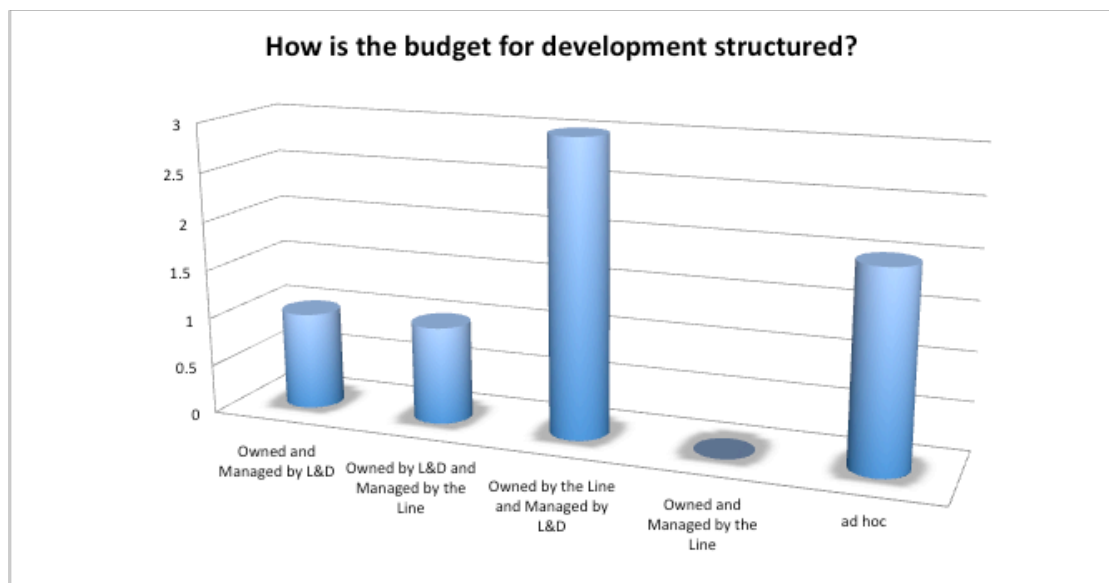
Most respondents commented that this would be some form of 'ideal' but that both organisational and learning culture did not work to make this possible. A notable sub-theme was the number of people resistant to the concept, seeing this as part of the L&D decision path and should not be left to chance (even though their own returns were poor).

How is the L&D function evaluated?

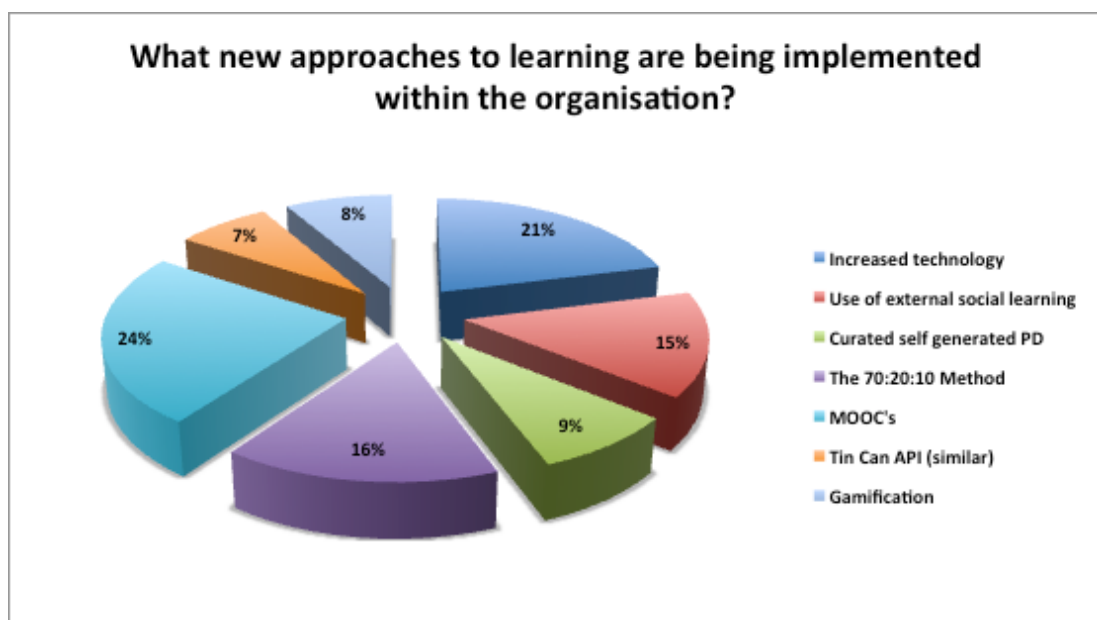


The figures in the chart represent percentage responses. At the heart of the problem is that L&D mostly operates within the unevaluated and less exposed area of HR. Many L&D people would prefer to be evaluated more, as they are confident they create and provide value; an interesting

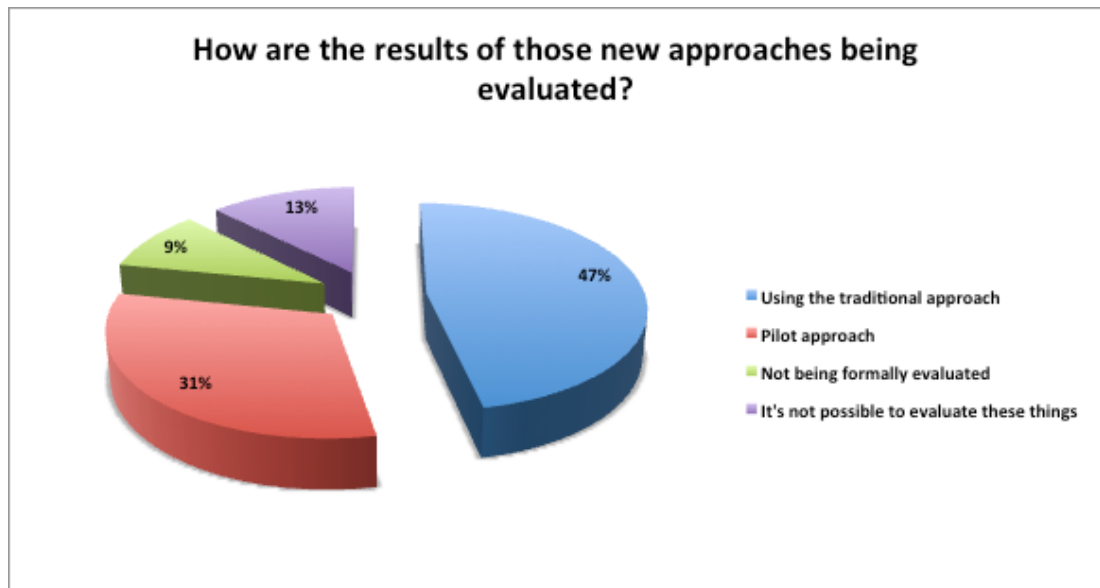
belief, given some of the internal process issues highlighted in this research.



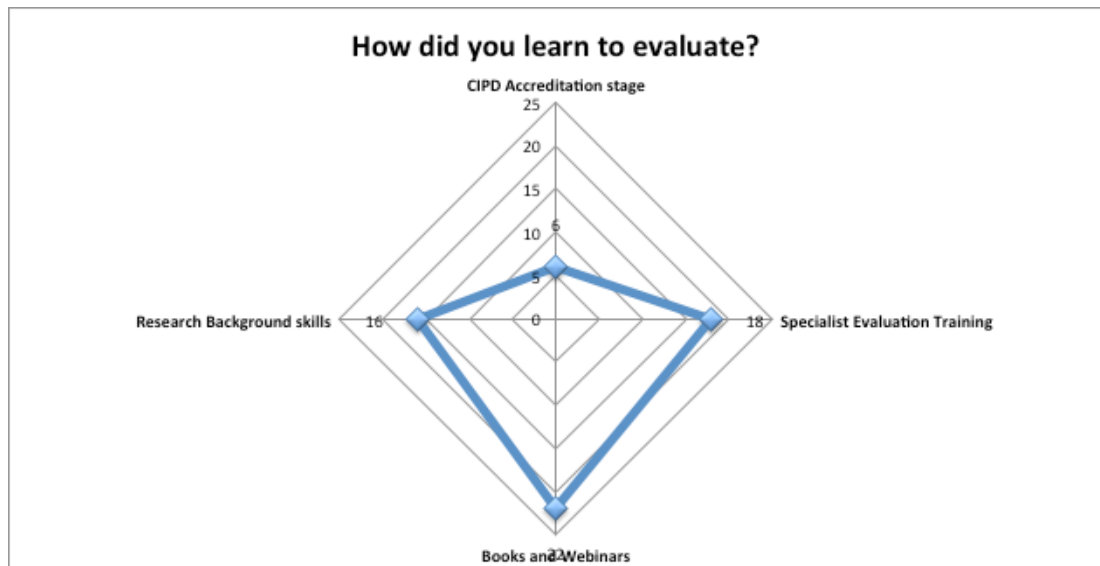
Again, the level of budget transparency can impede the process of accountability. One wonders, if evaluation were successful and robust, whether L&D would increase credibility and trust in adding value to the budget. This result reflects the Griffin (2011) concerns around L&D functional credibility.



Clearly, there is more innovation in this area. This is seen by L&D as what their function is really about: creating the opportunity and tools to learn – rather than the creation of evaluation processes that show 'how well' the learning has worked. However, more technology does not mean increased use of innovation or social media, and this is a lost opportunity, given the nature of change in the social use of technology, and probably simply represents more e-Learning.



Given previous responses, involving tools and processes, it would be difficult for innovation to exist here. However, what is interesting is that the IT function often evaluates a pilot IT approach using some of the techniques outlined in the review of literature. Perhaps, as training becomes more commoditised because of the increased use of technology it may be more prudent to transfer accountability for learning to that function?



Perhaps this reflects some of the thinking of Griffin (2011), in the lack of standardised approaches or the lack of professional expertise in evaluation, other than the courses delivered by those with the conventional methods or IT systems at their core.

4.3.1 Conclusion

The findings in this Phase of research can be seen to be disappointing, as well as exciting, in equal measure. Disappointing as they show the lack of innovation, concern and desire to evaluate – and exciting at the size of the opportunity in the L&D practitioner base for a solution that can solve the disconnect between the desire to have the data to drive better decisions in a simple and non-resource intensive way.

Given this is a group of people probably more interested in evaluation than the norm, as they are already evaluating, I rather expected better results, but the strong underlying themes within the responses highlight the desire to be better at the subject. As one person pointed out “if the organisation entrusts us with £100,000 to develop and deliver a programme, they should be sure of the value – otherwise they should spend it elsewhere”!

4.4 Phase Three — Case Study Outputs

4.4.1 Phase Three — Introduction

Having carried out two surveys and analysed the results, considered the literature review, and continued to carry out evaluation within my own client base, the new evaluation framework idea had been created, and was becoming constructed into a working framework with a series of tools.

Therefore, the case study process was designed to investigate:

- Whether the framework was robust – by comparing results to a retrospective set of data
- Whether the framework was accurate – by testing the approach against a third party 'gold standard' set of outcomes
- Whether the approach was credible – by assessing the outputs against the target expectations of the L&D practitioner group
- Whether the framework was cost effective – by assessing the accuracy against the relative costs of both the new approach and the third party 'gold standard' process

This phase of the research involved a number of process steps:

Steps	Outline	Output
One	Creation of the framework and its subsequent testing with a group of experts	A number of challenges were surfaced and are discussed within the methods' section A group contribution to the levels of acceptable forecasting accuracy (this output is shown alongside the requirements of the L&D clients)
Two	A Retrospective Analysis comparing previously collected data with the new framework	Comparison of the accuracy of the framework with the previously collected ROI scores using another method by another person
	The creation of forecasts for a range of programmes and courses for each of the organisations.	The establishment of the targets deemed acceptable in terms of accuracy by each L&D client Comparison of the accuracy of the framework against the method

		operated by a third party and compared to the respective targets identified by each L&D client The comparison outlining the accuracy by organisation and course
	The creation of a cost profile of the different approaches	An illustration of the days consumed by organisation and course to highlight resource savings.

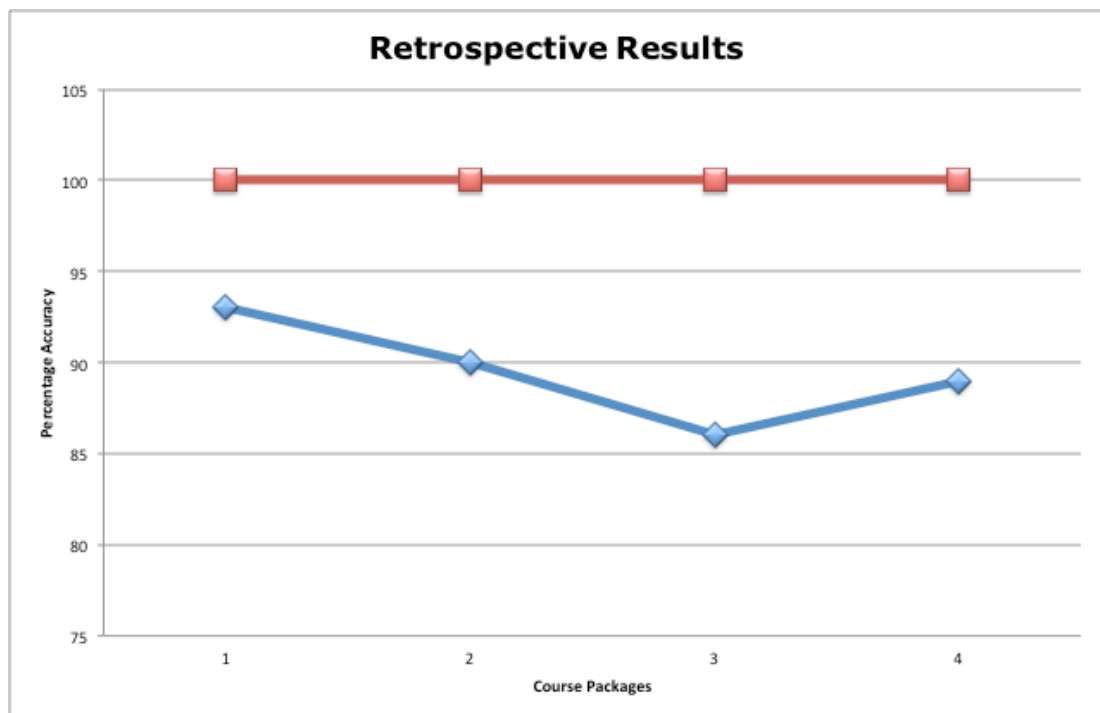
As outlined in the table above, the outputs from Stage Two will be shown as well as one chart from Step One. This illustrates the point that Stage One took place at the formulation of the framework concept and the outputs were more about thinking and challenge. However, one chart does appear in the next section and will be discussed at the appropriate point.

4.4.2 Results from the Retrospective Test

Courses were selected from an existing client, where I had no knowledge of the result, as the evaluation had been conducted by a colleague; therefore, the results from the new framework would be uncontaminated by prior knowledge. Twenty seven courses were used to test the new framework because, as Armstrong (2001) stated if the process did not work retrospectively it could not be relied upon to generate a secure forecast on unknown data. The courses evaluated were as follows:

Course No	Description
1	A Two Day leadership programme for Middle Managers
2	An Executive Coaching Programme for Directors
3	Five Mandatory Courses
4	Twenty, one-day skills-based courses for managers and non managers

As a reminder, the red line is the benchmark 100 per cent. score against which the blue (actual) line represents the actual forecast accuracy. The broad results were as follows:



To recap on the process used, each ROI output was compared as a percentage of the 'Gold Standard' process used in the previously collected data. So if the ROI figure generated was 300 per cent, and the framework produced a figure of 270 per cent ROI, then that was a 90 per cent accuracy measure. That 300 per cent figure was then converted to be 100 per cent more clearly to show the percentage accuracy of the framework. As the courses were all different, with extremely diverse returns, the 90 per cent figure became the accuracy percentage reported.

In a sense, I was not concerned with the actual ROI figure generated, only whether it was sufficiently close enough in accuracy to the original figure as a percentage. Whilst not strictly necessary at this point, this process was important to trial, as when comparing forecasts later with much more information, differences in courses and ranges of ROI, the ability to have a straightforward measurement point for practitioners would be a key feature of the framework.

Also, at this time, I was using the level of forecasting accuracy generated as a result of the expert group, which was averaged at 85 per cent, and adhering to the standard that the framework should never be more optimistic than the 'Gold Standard' data. This had been the issue with Utility Theory and had led to decreased credibility with 'line managers', as reported in the Literature Review. Therefore I was determined to learn from their experience, and not exceed the traditionally sourced data, which became the benchmark ROI score. This restriction placed a clear and rigorous standard and ensured a more realistic and pragmatic view of the courses and programmes evaluated using the framework.

Within the range of courses in the section entitled Course 4, one of the forecasts returned accuracy forecast of only 59 per cent, and three returned accuracy forecasts between 76 per cent and 82 per cent. A number of remedial actions were taken to improve the accuracy. In the lowest accuracy case, this had been due to an input error and, for the others, some fine-tuning took place in the adjustment of the internal software. Whilst the initial results had met initial expectations, re-running

the process improved the overall result, and helped to build confidence in the approach.

Having operated a broadly successful first trial, attention now moved to using the framework in real time, and against a third party operating the 'Gold Standard' Phillips approach (or the Success Case Method, if deemed to be more appropriate by the third party evaluator – that decision was theirs and theirs alone).

4.4.3 Stage Two – Results from Actual Forecasts

In order to assist with the numbering contained within the charts, it may be useful to outline some of the numbers and letter used to avoid unnecessary confusion. Broadly, organisations are represented by numbers and courses by letters.

For example, in the methods section, each company has a description within a brief pen portrait that can be summarised as follows:

Company Number	Description
1	International Pharmaceuticals Company
2	UK Housing Association
3	UK Retail Group
4	European Finance Group
5	Not for Profit Regulator and Professional Association
6	International Technology Company

Each organisation decided on three courses (or programmes) that it wanted to have evaluated.

Organisation	Course	Description of Work
1,5,6	A	Valuation of a whole budget
2	B	Valuation of 30 fundamental job skills programmes
3	C	Valuation of a number (four) of Leadership & Management programmes piloted within a division having previously been run in other divisions
5	D	Valuation of an eight module soft skills/behavioural personal development programme
2,4	E	Valuation of mandatory rolling cycle of programmes (six) using blended learning
2,6	F	Valuation of series (eight) of technical programmes each one day long
3,4	G	Valuation of mandatory health and safety programmes (four)
1,4	H	Valuation of induction programmes
1,6	I	Valuation of an accredited Leadership programme
3	J	Valuation of programmes to third party stakeholders
5	K	Valuation of a rolling programme of CPD
1	L	Valuation of large organisational initiative (e.g. Customer Service KPI change)

The same information portrayed slightly differently:

Organisation	Course	Description of Work
1	A, L, H	Valuation of a whole budget - Valuation of large organisational initiative (e.g. Customer service KPI change) - Valuation of induction programmes
2	B, E, F	Valuation of 30 fundamental job skills programmes – these were identical programmes, repeated a number of times with different groups across the budget cycle - Valuation of mandatory rolling cycle of programmes using blended learning - Valuation of series of technical programmes each one day long
3	C, G, J	Valuation of a number of Leadership & Management programmes piloted within a division having previously been run in other divisions - Valuation of mandatory health and safety programmes - Valuation of programmes to third party stakeholders
4	D, F, H	Valuation of an eight module soft skills/behavioural personal development programme - Valuation of series of technical programmes each one day long - Valuation of mandatory health and safety programmes - Valuation of induction programmes
5	E, A, K	Valuation of mandatory rolling cycle of programmes using blended learning - Valuation of a whole budget - Valuation of a rolling programme of CPD
6	A, E, I	Valuation of a whole budget - Valuation of mandatory rolling cycle of programmes using blended learning - Valuation of an accredited Leadership programme

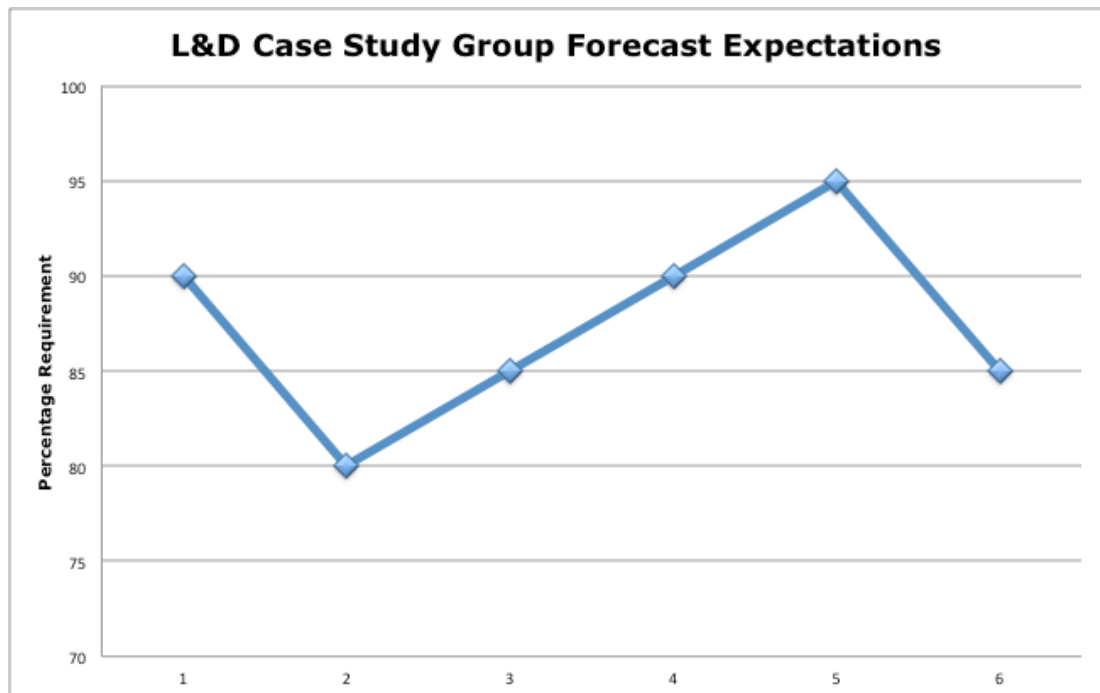
Having decided what each company needed, an internal evaluation project plan was created, and the third party evaluator was briefed. They then conducted their own process independently of the framework. No sight of each other's data was shared, until the final results from the third party evaluator were delivered.

The first process to be decided was to ascertain the level of forecast accuracy that each L&D client would expect to satisfy their own definition of credibility, and to build their trust in the framework and the overall approach. Initially, I had collected the opinions of the expert group, during the framework creation stage, and these accuracy expectations *for each individual* were as follows:



The range of results (from 60 per cent accuracy to 93 per cent - without exceeding the benchmark data, or the 100 per cent figure), was surprising, and reflected the relative levels of scepticism of the group. Bearing in mind that many of this group were fellow expert evaluators, the level of expectation was lower than I had expected. However, it is fair to report that the framework was still in a conceptual state at the time of discussion and that some colleagues may become more confident in the framework as its development progressed.

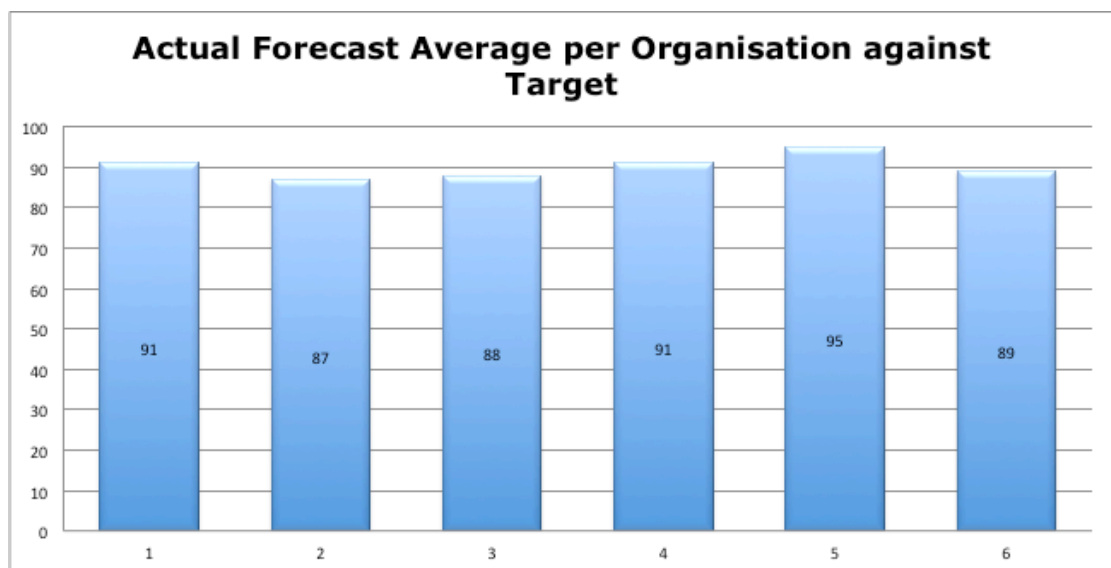
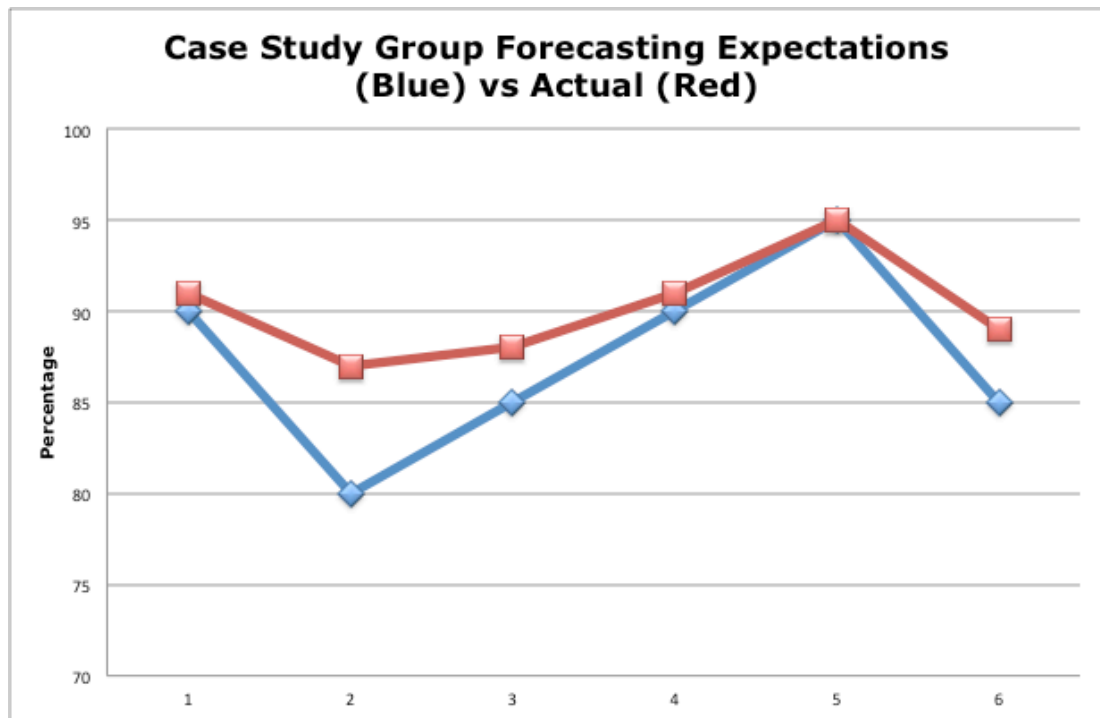
During the actual case study research and having briefed the L&D clients thoroughly, they decided their own levels, which were substantially more challenging than the expert group. They were individually as follows:

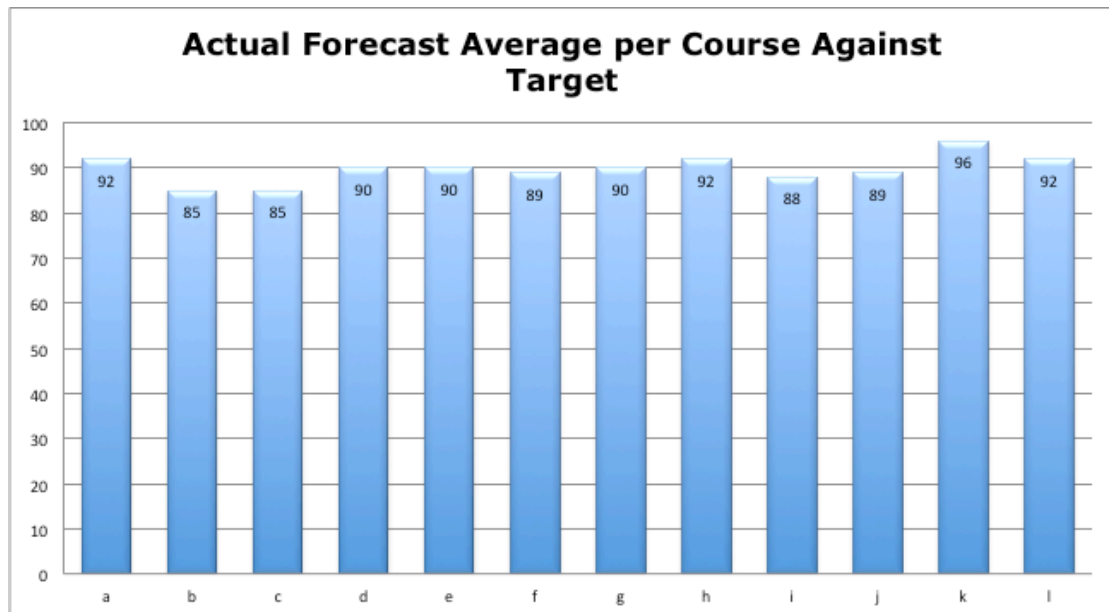


As one person said "put your money where your mouth is" and then created the expectation of a 95 per cent accuracy rate as something they could begin to believe in. This was challenging and left very little room for error if not to exceed the 'Gold Standard' Phillips benchmark data.

In the event, the framework calculations were calculated and were collected and stored. The third party conducted their summative process, and the results were compared. Again, the process outlined above of side-lining the actual ROI numbers, and simply creating percentage accuracy comparisons, was carried out.

The following chart shows the actual, average accuracy score from the new framework (red), per course, relative to the accuracy expectation (blue), of each of the L&D clients. One of the most peculiar sensations of this whole DProf journey, the money spent, the hours of effort, and the sweating of brow, came to fruition in this single, simple chart.



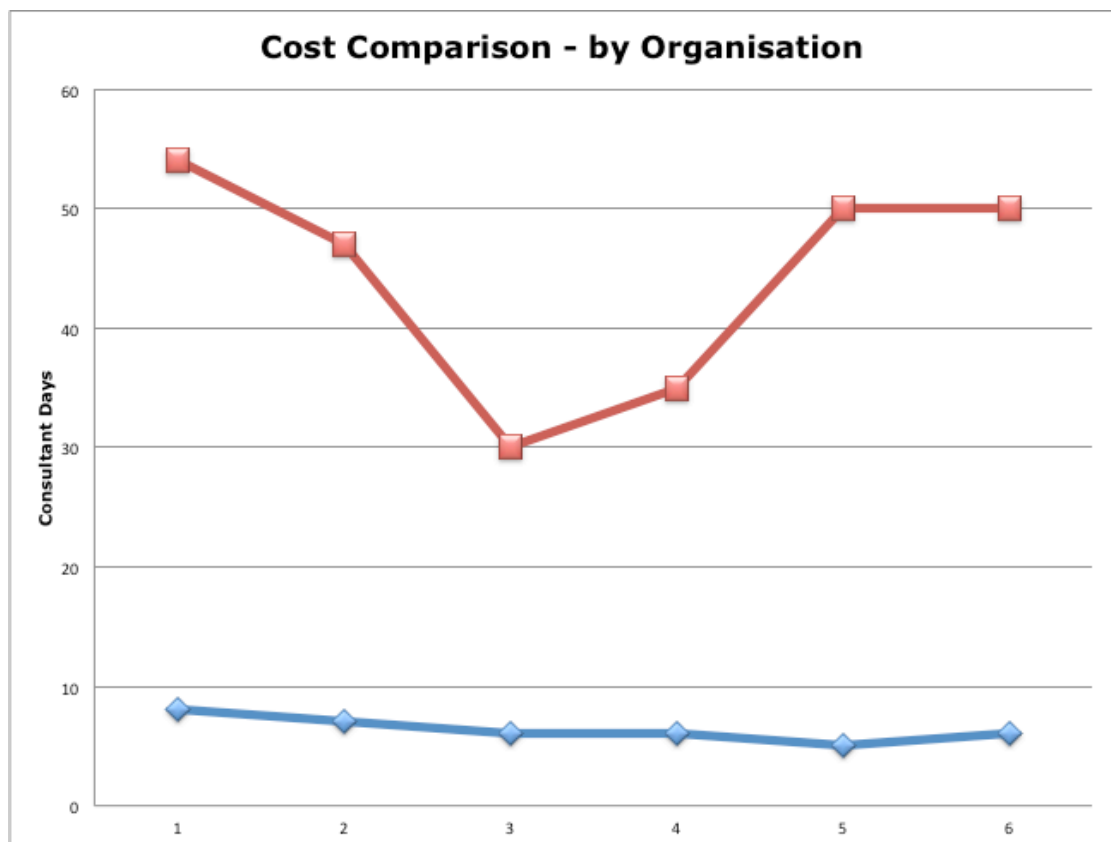
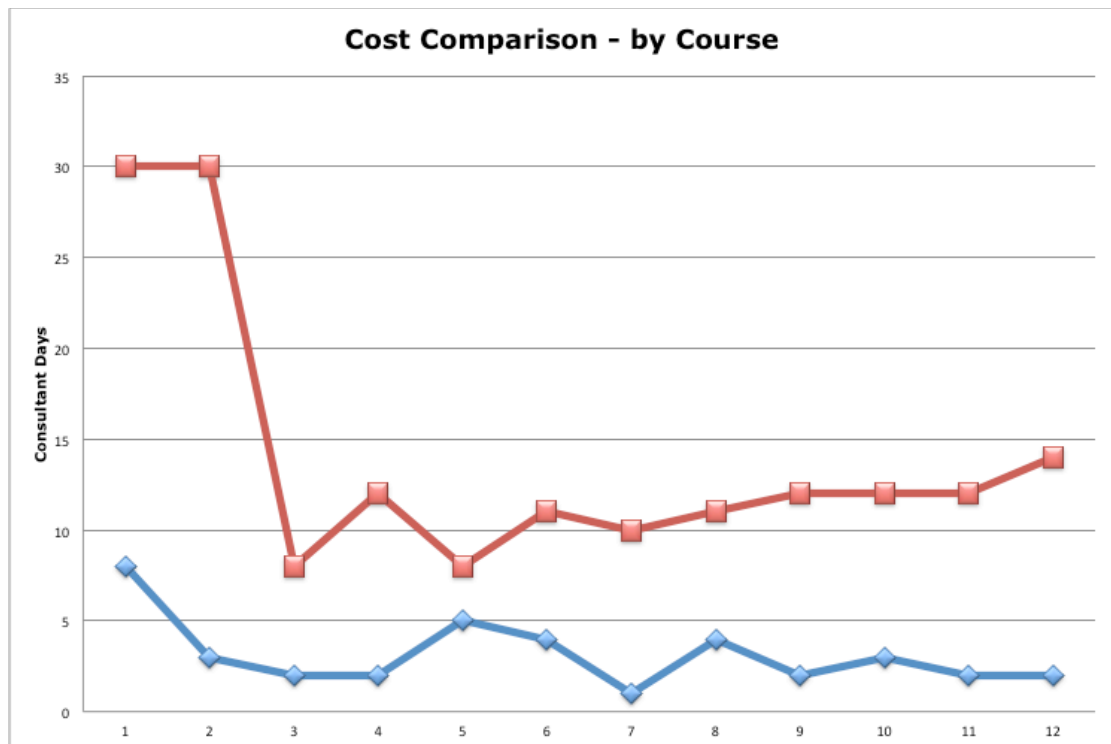


Whilst the results were extremely pleasing, the next, and possibly most important, challenge in potential adoption of the framework by the L&D function, would be whether the framework is cost effective. In other words, would the level of loss of precision in the use of the framework be justified in reduced cost and less L&D resource consumption and/or load in the organisation.

4.4.4 Initial comparison of transaction costs

The decision was taken to compare each of the approaches in cost terms, by comparing the numbers of days taken by the two approaches. A total cost comparison made little sense because of the nature of differences in 'daily rate fees', and the relative numbers of people involved in the process. Also, in order to determine the 'load' on the organisation, time became the key measure. There are four main parties in the process: the evaluator; delegates, their managers; as well as the L&D client. It is important to state at this point, that the framework involves no delegate or manager load as the process only involves the evaluator and the L&D function. This is important in order to reduce the effort, and problems of data collection, as well as the continued problems with 'surveyitis', reported by organisations that encounter resistance in constantly surveying and polling employees.

In both of the following charts, the red line represents the Summative 'gold standard' process compared with the use of the framework.



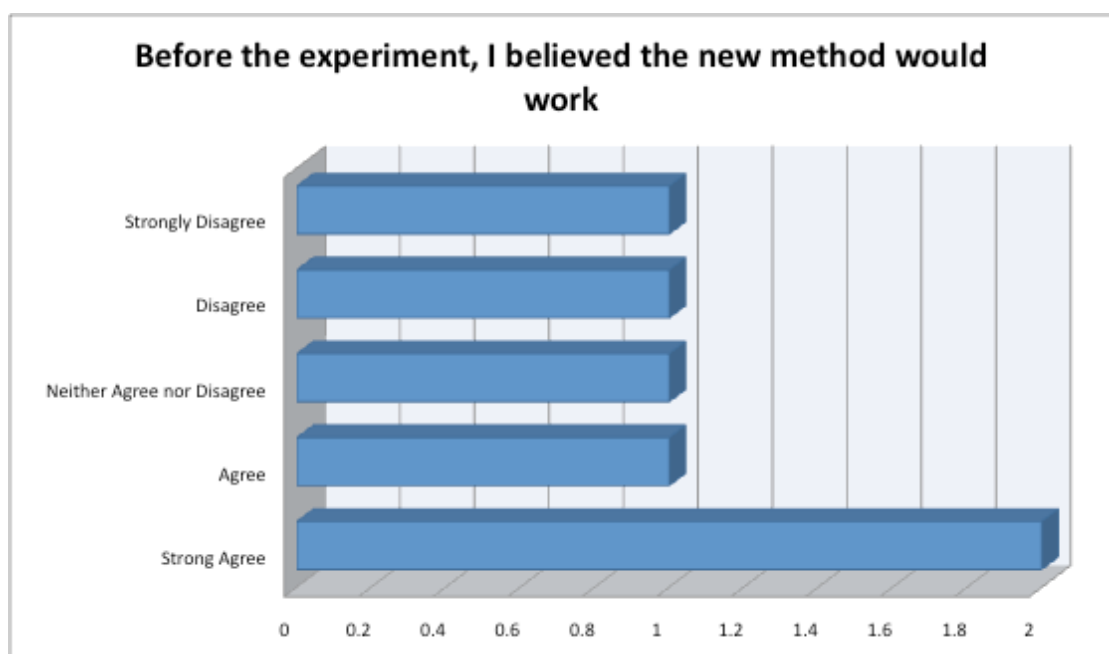
Again, the comparison is stark. The time, effort, cost and load on the organisation is substantially reduced in using this framework.

In terms of numbers, the whole evaluation process took 38 days for the framework, and 170 days for the third party process all the courses evaluated. At an average cost of £1000 per day, this saving becomes considerable. Given the level of accuracy achieved, it is proposed that this framework could offer the L&D function a solution where there is actually little trade off in accuracy in calculating ROI, for a substantial saving in cost.

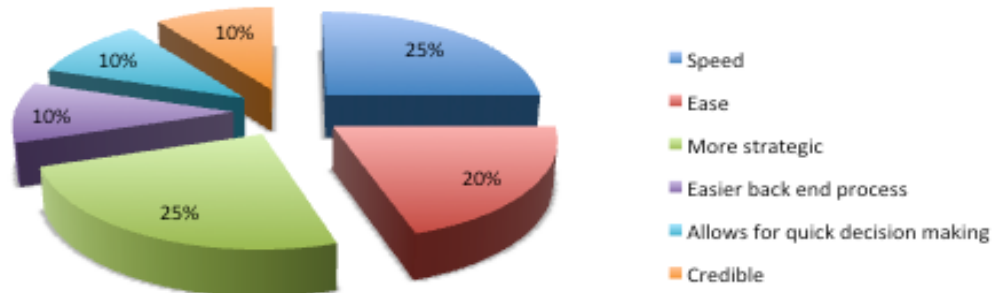
4.5 Phase Four — Overview and Results

Having completed Phase Three and analysed the data collected, it became appropriate to carry out a short opinion survey of those who had used the new approach.

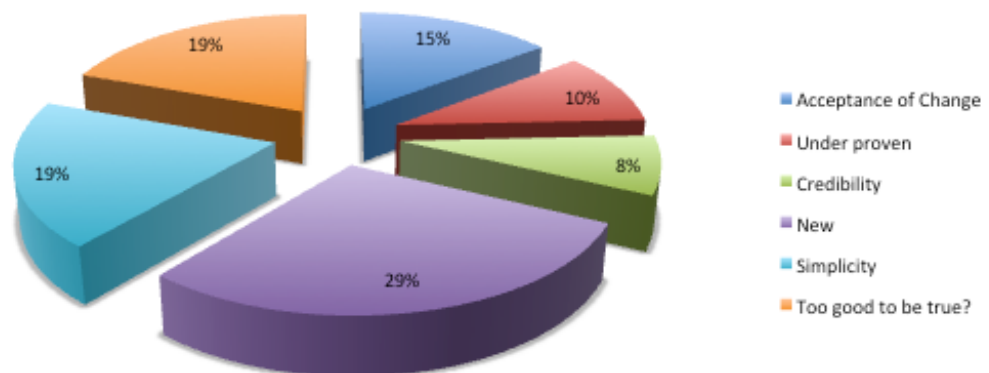
This consisted of a short online survey. The charts need little or no commentary.



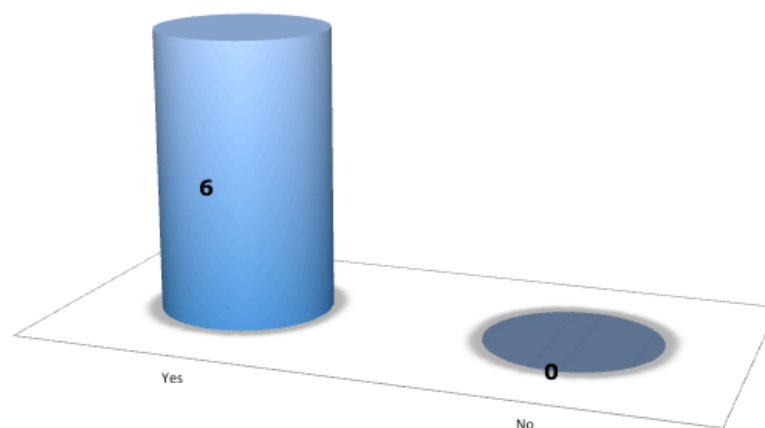
What do you think are the benefits of this new approach?



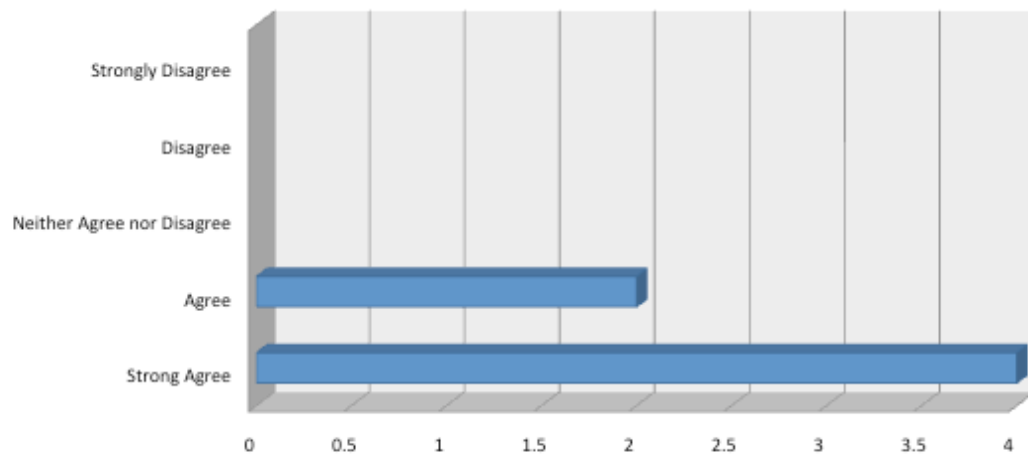
What do you think are the risks of this new approach?



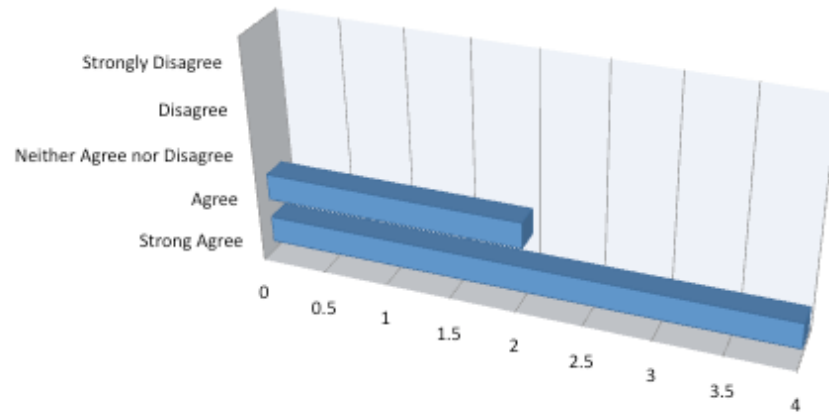
Do you believe that the cost savings are an acceptable trade off for the reduced precision of the framework?



I now trust this approach



We will carry on with this new approach



4.5.1 Stage Four – Reflections

The response from the case study group created a number of useful observations and comments, that can be used as part of the continued development of this process, as well as to stiffen the robustness of the approach. They include:

- Link more of the academic background into the approach to give it greater legitimacy
- Create a more pragmatic tool-kit for its deployment with greater transparency of the inner working within the concept
- Consider a different weighting mechanism in considering the variables
- Consider a more stringent 'test' with a shorter forecasting period
- Build the ability to forecast longer into the future
- Build a better reporting suite
- Try comparing the approach with other evaluation methods

Most of the case group recipients stressed the point that the simplicity and obviousness of the approach was both its strength and weakness (somewhat like the Kirkpatrick idea.)

As a whole, the group thought the approach could replace Kirkpatrick, but only if it 'caught fire' somehow. A few suggested that the line management population should be trained in its use so that they could use it to hold L&D to account.

4.6 Summary of Research Activity

The research findings outlined in the document span Four Phases of research. They consisted of: an opinion survey of a purposive sample of practitioners, a more targeted purposive sample of practitioner involved in evaluation, and a case study with a number of stapes, and a final opinion survey of those taking part in the Case Study. This chapter has laid out with some commentary the key findings; more discussion of the research follows in the next chapter.

Chapter 5: Research Project Discussion / Interpretation

5.1 Introduction

This chapter discusses and interprets the findings and meanings from the research, relative to the literature review, in order to situate the relevance and applicability of the findings within the wider literature. From this, overall conclusions will be generated from which a number of recommendations for both practice and further research are made.

The evaluation literature contains a broad range of themes, schools and thought and it overlaps, in many cases, with the technical aspects of 'learning,' 'performance', and 'behaviour', as well as associated areas of practice, including social research and marketing, with areas of commonality and shared frameworks and perspectives, from which mutual learning has been drawn. This chapter is structured as follows:

- A reflection on the key themes from the literature review
- An interpretation of findings from each of the research phases
- Some suggestion for future research

5.2 The Literature Review — Reflection

The literature review found a range of insights, both through structured reading, as well as the association of reading from wider and less formal areas. The review initially considered the body of work most closely associated to the evaluation of L&D which was grouped to stimulate insight and to give a fresh perspective. The review then branched into areas of research necessary for the construction of a new framework.

Some of the key insights include:

- A considerable body of the literature review reinforced the primacy of the Kirkpatrick method, despite substantial evidence and research demonstrating issues and problems. These ranged from a critique of its conceptual standpoint, to detailed criticism

of each of the internal levels, as well as the tools and methods within each section of the framework. From Holton (1996), to Griffin (2011), including Cascio and Boudreau (2011), as well as Kearns (2005), Spitzer (1984), Phillips (1996) and Kraiger (1993), the most respected evaluation authorities of their day were dismissive of the concept and its application. Holton (1996), was particularly critical of the framework and those L&D functions still adhering to it. However, one of the interesting reflections was the degree to which evaluation frameworks sought to build their own legitimacy by building fixes and solutions for the various flaws in the Kirkpatrick framework and have, themselves, become successful, particularly Phillips (1991).

- The prevailing attitude within the practitioner base, to evaluation suggested a lack of focus, approach, and justification of value, despite the fact that a single framework was so well established in the practitioner base. The level of time and effort required to make the Kirkpatrick framework deliver legitimate results was less beneficial than concentrating on other areas of L&D that were possibly more interesting and less time intensive. The inability to 'do Kirkpatrick', rather than 'do evaluation', is the issue in hand that must be solved.
- There is a tangible disconnect between the needs of practice and theory and, despite a wide field, very few ideas, other than Kirkpatrick and Philips, had gained acceptance. The various schools of thought often focused on a theoretical perspective of evaluation not linked to practice, constantly generating 'solutions' that were rarely adopted by the majority of the L&D practitioners, and it become increasingly difficult to disagree with Griffin (2011), that the function had decreasing degrees of credibility, as there was no cohesive thinking or approach that held together both academic and practice-based evaluation.

- The level of innovation, outside the narrow confines of the evaluation literature, was in stark contrast to those shaped by the Kirkpatrick paradigm, and offered new perspectives, approaches, and concepts, upon which a new evaluation framework could be based or constructed.
- The opportunity presented by the ideas shown in the associated school of learning embedding that needed to be explored in more depth to create mutual opportunities for practice.

It is recognised that any literature review will be affected by the reflexive approach operated by the practitioner, and that others may draw different or fresh inferences from the same body of literature. However, the conclusions drawn from the review drove a practical approach to the research, by establishing a number of questions that could create a focus for the research, as well as surface findings relevant to the creation of a case study, to investigate the practical application of a new framework. The areas of research were split into two primary, and five secondary, questions.

5.2.1 Primary Questions

- How can a new evaluation framework, built upon Decision Theory principles, add value and allow L&D to prove value in a cost effective way?
- How can the competing drivers for the need for change and the need for credibility in evaluation be resolved in a function under pressure?

A range of sub-questions were refined as follows:

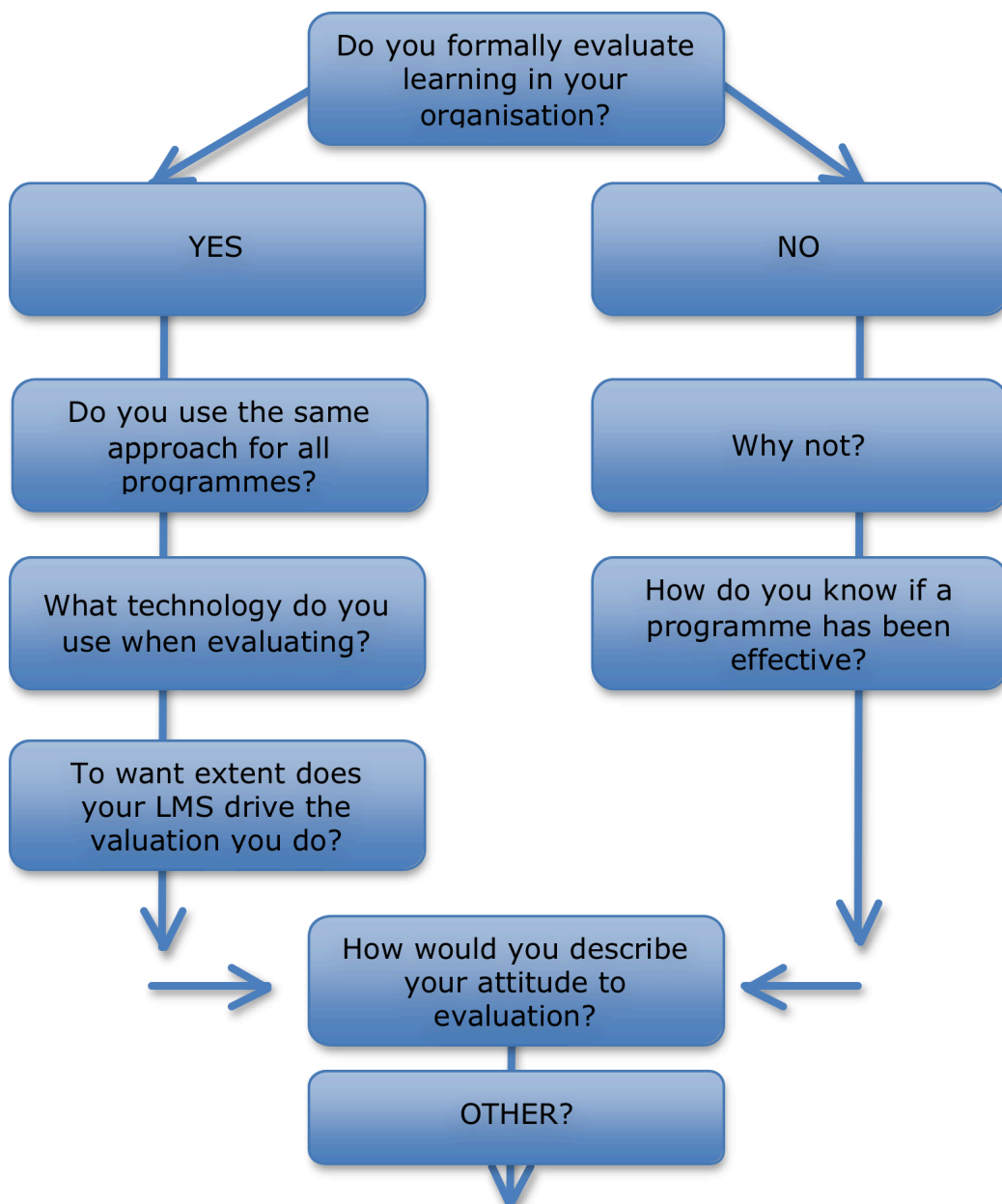
- What new or existing attitudes exist in the practitioner base that continues to drive or restrict attempts to evaluate?
- Does the outcome required by stakeholder create some of the confusion at the heart of the process?
- What part does credibility and resource effectiveness have to play in any solution?

- Does the longevity of the Kirkpatrick framework represent excellence, apathy or a lack of choice?
- Should the process of 'what is possible in technology' create the process of evaluation that organisations buy into?

5.3 Phase One – Attitude Survey

Following the literature review, without a clear idea of any solution to the problem, a pragmatic view was taken to conduct a brief and focused review of attitudes within a purposive sample of practitioners in the field. This helped guide a new framework, by identifying the enablers and barriers to a new approach, as well as to highlight the interaction of theory and practice at that time. It also served to enrich some of the practitioner attitudes, regarding reasons for the lack of good evaluation.

The structure of the research was as follows:



The most significant themes here were answered by the total population, and it became apparent that the concept, rhetoric, and practice of evaluation are different across the sample of the L&D community. People appeared to break down into groups of:

- Those who knew what they were doing, saw the value and had expertise
- Those who carried out some basic evaluation as a means of 'doing something' – this group was sub-divided into those who were satisfied with that level of activity, and those who believed they should have done 'more' (assuming they knew what 'more' was)
- Those who did little or no evaluation – this group were subdivided into those who were happy with that approach (no external pressure to do more), and those who wanted or would have liked to have done 'more' as above. One of the issues across the groups was the desire to focus more on areas of perceived interests and strengths (the new advances in learning and learning design), rather than an area of weaknesses, i.e. data, analysis and reporting.

Irrespective of which group the person belonged to, the majority stressed that evaluation was important – even though many of them could not do it, did not want/need to do it, or did not know how. The fact that so many recognised its importance, yet did not carry out effective evaluation, began to suggest that the problem was a result of the current solution, rather than the subject itself.

The initial findings showed a substantial number of organisations that carried out little or no evaluation. The CIPD (2015) report also found there were still 14 per cent of organisations surveyed at that time that carried out no evaluation at all. The findings are more extreme in this Phase One research, as a 58 per cent negative response may reflect the sample used or the process adopted, in creating a forced choice. The other challenge in the data was that, within the 'No' group, appeared a

range of informal evaluation techniques that showed some evidence of evaluation; however, it served to reinforce the contention made by Griffin (2011), that the lack of robust practice could undermine both the legitimacy of the function, and the credibility in the reporting of results.

To some extent, there was no real surprise in this result. What was more remarkable was the finding about attitudes to the whole subject area. The prevailing attitude within the group tended to be 'negative' or 'very negative'. This was surprising, given the nature of the group, which contained only those who had self-selected as being interested in the subject, yet were mostly negative about it. This may reflect the view of Hashim (2001), regarding the motives for evaluation, or a more pragmatic view, based on the commentary in the 'Other' questions, where a range of opinions broadly fell into two areas. Those who had no requirement to produce formal evaluation and those who simply did not have the time to do any. These two drivers were irrespective of the functional desire to carry out evaluation.

Similar attitudes were also reported by Holton (1996), and Phillips (1995), where both had found that the failure of the function to evaluate had drivers in both external demand and the 'time vs. value' argument. Their solutions to those problems were different. Holton moved in the direction of improving L&D process, through Learning Transfer, and Phillips in the creation of a more robust and financial legitimate ROI framework, to generate greater legitimacy. To find the attitudes still so prevalent implies that those solutions have not solved the underlying problems; indeed, it could be argued that the Philips' approach, by making evaluation more time consuming, has served to reinforce some of the attitudes. A very small percentage of comments in the research pointed to the attitude that, if the C-suite leaders and (particularly) if the Finance Director, do not believe that L&D can be measured, this removes the need altogether. This view may represent an interesting angle of research activity for others.

The awareness of the Kirkpatrick framework was almost total, both in terms of the levels of understanding of the concept, and its inner levels, and also reflected the increased awareness of the Phillips (1996) framework and the jargonised use of 'Level 5'. That awareness, and subsequent use of the Kirkpatrick framework for the majority of courses or programmes, represents a challenge to the various proponents, who advocate different approaches, given the needs of the stakeholders, either undergoing or commissioning, the training. Kirkpatrick and Phillips particularly represent the needs of the organisation, as the primary stakeholder for training and opportunities, to utilise the work of Geertshuis et al. (2002), to reflect learner-driven evaluation, and of Kraiger et al. (1993), to build a learner-focused culture with a 'partnership' in outcomes for both learner and organisation, however, such needs are not reflected in these findings.

Many in the field of evaluation research have relied on technology to produce results from summative evaluation. For example, the use of SPSS, and other analytical tools, as well as coding software solutions for content analysis and other qualitative methods, have been at the heart of frameworks by Fitz-enz (1994), Kraiger (1993), and Brinkerhoff (2003). However, the prevailing technology in use in L&D (where technology exists at all), is the Learning Management System, a technology created to streamline process and reduce the 'cost per head' of training within the function. Scoutardis and Dyke (2007), found that inferior evaluation is often generated where the LMS is the engine for evaluation. The findings showed a high dependency in the use of the LMS both to manage the process and generate the outputs of evaluation; sometimes supported by third party evaluation reporting or analytics resources. This is a concerning trend for a number of reasons as, according to Tasca et al. (2010), the majority of LMS's use the Kirkpatrick method to deliver the results, and the use of LMS's produce a low yield of results. Also, a rapid review of the websites of LMS suppliers still shows their primary motivation is to use evaluation data to demonstrate their own success in process cost reduction.

At the other end of the spectrum, were the numbers of respondents using 'no' technology at all to enable or facilitate the evaluation process. Using a paper-based system would be a challenge, and there would be a significant load on a function, where manual systems could only be used. Few references to this lack of technology exist in the literature and insights in this area of process may be most likely to be derived from the marketing databases of the LMS providers. It was also a finding that the lack of formal technology existed in those organisations where 'informal' evaluation activity took place. Simple reporting of 'stories' and 'buzz' would be best served by manual systems, or perhaps to utilise influencing systems or formal communication processes for management reporting.

Finally, the findings showed the use of the external training company data for evaluation. This represents a real problem, both in terms of the legitimacy of any evaluation claims, as well as the potential ethical challenges for the external trainers. Given the Guerri and Vinante (2011), review of vested interests, and the need to reduce the level of influence of the external training provider, this represents either a flaw in concept, process, or competence for any internal L&D function, to allow anyone to 'mark their own work'. The rise of external regulation will also militate against this practice, as an examination of the requirements of many professional regulators (e.g. CQC, NHS) require evaluation to be separated from the service provider.

The findings regarding lack of time in the attitudes' section were reinforced in the reasons for the lack of formal evaluation. However, the subtlety of the finding was enhanced by a sub-theme within the perception of 'lack of value' in the result. If something is difficult to produce, and has little demand or legitimacy, then the opportunity cost for the process outweighs the effort involved. Phillips (1992), evidences the positive results for the ROI framework, where the legitimacy and credibility generated by the process, is shown to be worth the effort of those properly trained and effective in deploying the process.

Griffin (2011) worries that the approach to evaluation undermines both the credibility of the function as well, as the legitimacy of the reporting of results, at both the micro and macro levels. It could be argued that these findings may support that view and that the L&D function (supported by the CIPD), need to raise and improve both attitudes to, and skills of, evaluation.

5.3.1 *Reflection on this Phase*

The creation of an attitude survey reflected a pragmatic place to start the research process, and to begin to bring out some of the key themes. Whilst being informed by some key questions, they had to be flexed and adapted to reflect the learning I was experiencing as part of the wider process. A substantial gap in this phase was not to have examined some of the ideas around Learning Transfer at this stage. This may have led the research into a deeper examination of the overlap and interplay between the two concepts, and produced a different result. It is possible that some value or insight could be created for future researchers in examining this area, as the opportunity to produce a more active, interdependent process could produce a different practitioner tool-kit.

In hindsight, a more thorough inductive process, with more penetrating and insightful questions, may have discovered a wider range of themes, rather than such a heavy emphasis on a small number of themes. This adoption of a different range of questions may have led to a more focused second Phase of research, and may represent a lost opportunity to have gained insights from a larger sample size.

Whilst initially being disappointed with the findings from the Phase, and of the attitudes within the practitioner base, it did motivate me to decide that a solution could have a genuine impact, and led to more focused research, as well as wider reading. Phase Two was therefore constructed to examine the practice of those who were actively engaged in evaluation in the hope that practice had moved on in the gap between the two Phases of research.

5.4 Phase Two – Specialist Practitioner Research

This Phase of research followed a pause in the overall research process, because of a number of personal challenges, and therefore became intertwined with the innovation process used to create the new framework that would be researched as part of the Case Study in Phases Three and Four. This Phase of research brought some real insights for the evaluation literature, even though much of it did not directly translate into the new framework. Rather, the framework was created from some of the gaps in practice not contained in the findings.

The questions can be grouped into three main areas:

- L&D Evaluation practice
- Stakeholder Needs and Outputs
- L&D Process Improvement

5.4.1 L&D Evaluation Practice

It is encouraging to note that the findings showed substantially different outcomes from those of the more generic group, especially with regards to overall attitudes to the subject. The commentary included a range of pragmatic opinions, including one that illuminates the assertions by Holton (1992), and Fitz-enz (1994), that evaluators “need to be able to account for the money entrusted to them and show some definition of value”. It may also be the case that those with greater expertise have greater confidence in applying evaluation processes, and can then understand the benefits of their data, relative to the time and resources needed.

The range of development evaluated by many of these practitioners was encouraging, as was the increased use of evaluation approaches. Whilst there was still a heavy dependency on the Kirkpatrick framework, other frameworks were also reported, including the Brinkerhoff Success Case

Method, and specific techniques to help measure pilot programmes, often using control groups. Whilst control groups are popular with Kirkpatrick, there is much discussion of their usefulness elsewhere in the evaluation literature. Schmalenbach (2005), identifies that the scalability and speed of control groups can interfere with pragmatic evaluation, and create too long a 'lag' between implementation and the effectiveness of the learning process. Rae (1983), contends that control groups fail to control subjectivity in qualitative analysis and should not be used for evaluating programme 'roll-outs'. Scriven (1996), rejected control groups in favour of the use of external evaluators; however, there may be too few independent evaluators to make this a pragmatic solution. Finding control groups in use for pilots may reflect more informed understanding of social science processes through more academic qualifications, rather than practical evaluation development.

Another finding was the rise in the use of the Net Promoter Score ('NPS') as a strong sub-theme in evaluation metrics. Whilst not a method, more a simple score, finding this relatively modern concept being used in evaluation to link to 'satisfaction' could well be a sign that simplicity and organisational familiarity can lead to fairly rapid acceptance of a new concept. Whether NPS will gain traction is questioned by Hansen (2011) in Meta research, showing the increasing range of opinion opposing the use of NPS as too simplistic a solution to a complex problem.

The group was also using the Phillips' ROI method, linking it particularly to specific programmes, such as a review of an organisational KPI, or rolling programmes of Leadership Development, or Executive Coaching. It appears that the experience of Wills et al. (1996), in using the approach for a significant programme, where extensive spend is deployed is reflected in wider practice. The level of awareness in this group, regarding the use of approaches against need and spend, illustrated how targeted evaluation generated credible results and generated value, to build positive evaluation mindsets.

The motives for evaluation that had been a particular concern of Hashim (2010), appear to be less of an issue within this group, and this may explain some of the more targeted actions. Motives included the need to build continuous learning and knowledge acquisition, as well as to prove ROI. This is encouraging, as this reflects the contention that evaluation is part of a Decision Theory approach, rather than a disconnected review process. However, there is still a significant sub-theme to prove 'ROI' and this is troubling. Botchkarev and Andru (2011), contend that L&D use the term as a 'shorthand' and that functions often create the expectation they can generate a financial ROI figure without understanding what it is or how to create it. Also Kerns and Po (2012), maintain that using a ROI approach can destroy tangible value from 'soft skills' programmes, and Goldsmith and Sarno (2009), have a range of reasons why ROI is a poor measure overall.

However, within a more expert group of practitioners such as this, the use of the term could be assumed to be more correctly applied, although the point was made that ROI does not always have to refer to a financial measure, and can refer to Return on Expectations, or the aggregated perception of value. However, whilst the findings show some movement away from Kirkpatrick, it is still disappointing to see little, or no significant, use of newer measures or concepts, such as Utility analysis. Even the Kraiger et al. (1993), framework used as part of the Ford Academy appears not to have spread as widely as hoped or to have reached this group. The reliance on Kirkpatrick was often restricted to the 'Level 1' activity of measuring 'Reaction'. This is a concern as the literature showed wide distrust of this stage, including from Darby (2006), who worried about peer pressure affecting results; to Noe (1986), who dispute the correlation between satisfaction and learning and dislike that Kirkpatrick assumes this as part of Level 1.

Reports from Learning Skills Technology conferences and exhibitions as well as the CIPD (2013,2014), highlight the interest shown by L&D in the propagation of new methods of learning, and a range of technology and process ideas appeared within the findings. The number of self-directed

and curated learning experiences was interesting, as well as the rise of the newly popularised 70:20:10 concept of learning; potentially another new challenge for evaluators. However, it appeared that the innovation in evaluation has failed to keep pace in this area with the usual approaches being deployed to attempt to understand impact and value. Some effort had been directed towards using summative processes in Social Media - but with little success. This is a lost opportunity, given the work of Furio (2013), in developing iPhone apps to evaluate a programme, as well as the work of Landers and Callan (2012), in evaluating the impact of learning within a 'Virtual World' environment.

The lack of pace from the translation of academic knowledge into the practitioner base appears to be especially slow. This was an inadvertent finding, as many evaluation solutions exist for the areas of innovation shown above, but there appeared to be insufficient visibility for the practitioner base. One of the challenges propagating the status quo was the nature of professional development for evaluators. Findings in the research showed that the majority of development of evaluation practitioners appeared to be from the specific methods and processes associated with a particular methodology, most regularly the Kirkpatrick and Phillips' methods, which propagate their own tools and vested interests. Whilst the degree of self-driven learning was encouraging, perhaps greater signposting of useful content could help the practitioner base, and this is unlikely to be forthcoming from the professional body, with its own reliance on the Kirkpatrick framework.

5.4.2 Stakeholder Needs and Outputs

The findings from this section aimed to illuminate two factors. First, the degree to which stakeholder needs were a part of this evaluation process and therefore informing decisions, as well the value and utility of outcome reporting, and the impact on L&D functional credibility.

The findings about reporting were encouraging. They showed that data were being used to inform a wider discussion about improvement. This

could have referred to a supplier, or the quality of learning delivery overall, and this data informed internal meetings and served to illustrate the informal use of Decision Theory in the more sophisticated L&D functions. In many ways the Dessinger-Mosely (2006), framework was being loosely applied with a clear aspiration to compare results with intentions, and to show both the value of the outcome and the effectiveness of the evaluation itself. However, it is unlikely that the functions were consciously operating this framework as the findings showed measurements in use were less rigorous than that particular framework would require.

Within the reporting was evidence of the creation of data to be used to populate an HR dashboard. The commentary around this area reflected the need for HR to receive more process-driven metrics, or cost control data, rather than value creation outputs. This seems an approach reinforcing the Ulrich (2002), service delivery concept, and a far cry from the aims and ideals of the Human Capital school of thought, where Plant et al. (1992), linked the purpose of HR to the realisation of the value of 'people assets', which are shown to 'drive' organisational performance and value. Given the quality of decisions shown in the findings made by L&D, as well as the data they are capable of surfacing, perhaps being located within the HR department could be argued is a problem for both the credibility and perception of value in the L&D function.

The findings show that data were being used for a range of reasons. They also showed that the ideas of Fitz-enz (1988), in using data to prove value; of Flamholtz (1985), to use data to prove improvement and justification, as well as Cascio (1991), in using data to justify budget and legitimacy were being applied. This illustrated that Decision Theory was being informally operated, and that decisions were linked to both inputs and outputs across the whole L&D process. Whilst the findings show there was work to be done in the creation of reporting tool-kits, and quantification of value, these findings may counter some of the ideas of Griffin (2011), who worried that the lack of a robust approach to

reporting would detract from the perceived value of the function, and be at odds with some of the conclusions from Phase One.

One of the challenges for L&D is the way the function itself is managed and evaluated. This 'meta-evaluation' is a key part of the ideas of both Dessinger-Moseley (2006), and Passmore (2012). The findings showed this to be both a problem and an opportunity. The most significant financial measure was the use of the budget to manage the function. This implied that the function was simply expected to deliver activity within a specific cost framework. However, where the budget was located could be a problem for both the organisations and L&D, as there was a mix of ownership and management of the budget between L&D, HR and the 'line'. It could be the case that using more common practices valued in the line management population (rather than the virtually unique evaluation frameworks), could build legitimacy for any outcomes proposed by L&D, other than simple budget adherence, for example, the use of KPIs was already in existence, and linked more strongly this area of shared understanding and opportunity.

One of the most troubling sub-themes in the findings is the extent to which L&D function was not evaluated at all. This reflects some of the concerns from Griffin (2011), about vulnerability to external forces and legitimacy, but this strikes me as being significant given that the majority of those respondents who carried out evaluation, but were not being held to account for their results. In every case, those L&D departments were situated within an HR function, and some reinforcement of the argument about the role of HR posed earlier may be supported by this finding.

5.4.3 L&D Process Improvement

The findings showed that the processes of commissioning appear to be more diverse than those shown in CIPD reports (2015), and this is significant as the CIPD contend that as nearly 50 per cent of training needs lead to bespoke outcomes, then this stage is increasingly

important for evaluation in ensuring the correct outcomes and that decisions are generated at this point. Anderson (2007), points out that learning should have pre-agreed learning objectives to meet both the needs of the sponsor and organisation. The reliance on pre-course objectives and lack of tangible actions, learning points and targets for the training outcomes, is a problem here as Basareb (2007), and Griffin (2011), contend that programme objectives are not specific enough to drive the actions required from the programme. Anscombe (1959), also contends in her work around intention that specific 'first steps' need to be established, rather than vague or broad aims or goals, to ensure that a learner has a greater chance to carry out post-course actions. Whilst the findings show that L&D factored evaluation into both the design and delivery of the programme, the most used Kirkpatrick framework offers little guidance in terms of content in these areas, other than the use of competency frameworks and 'reaction' to understand post course impacts.

The findings showed that the interest in factors such as confidence and engagement as a course output, and that was encouraging. Although some commentary existed about the need to move away from 'fluffy' concepts such as these, it is a frequent contention in evaluation practice that is the role of the evaluator to create measures for these areas, as they are important. Whilst Chiaburu and Lindsay (2008), contend that effective, individual learning 'triggers' need to exist to drive engagement, the work of Monroe (1985), and the diagnostics of Hasan et al. (2014), begin to highlight the links between 'Perceived Value' and Engagement and, given the rise of this measure on many HR scorecards, driven by the work of Gallup and others, it would seem remiss not to concentrate some effort in measuring this area, as the demand for a solution clearly exists.

One of the challenges, shown in the findings, reflects a concern in many areas of summative research, even in the wider social sciences; that of question response. The findings showed a continued problem in this area, with a significant reduction in post-course response. This is a concern for Casey (2006), who also linked the increase in subjectivity of responses,

as well as legitimacy of data from reduced responses in summative methods, particularly the Brinkerhoff (2003), Success Case Method, that relies on a purely summative and subjective process. Despite the range of methods shown in the findings to improve the process of response rates, it is the contention of the Learning Transfer field of literature that it has effective solutions in this area. However, a different approach would be to innovate totally, and move to the concepts espoused by Spitzer et al. (2005), and Basareb (2007), in moving to a purely predictive framework.

The Learning Transfer field improves the post-course implementation of learning through effective 'transfer of learning'. The findings showed that efforts existed to achieve this through a range of process actions, including post-course coaching, where the commentary enthusiastically advocated the use of coaching to reinforce training, particularly in regard to executive level or senior level management and leadership programmes. Another process action in the findings was the use of blended materials, to reinforce key aspects of the learning post course. This is a key process tactic advocated by the Transfer school, and brings the concept of 'sticky' learning to life, as recommended by Holton et al. (2000). One disappointing method, within the findings, was the use of post-course evaluation forms to create embedding. Given the low response rate of questionnaire returns, used independently of other embedding methods, it is difficult to see how this could add value. An encouraging trend was the use of some newer technologies to embed learning, for example, using gamification, competitions and online groups to foster post-course conversations and impact.

One of the challenges for Learning Transfer, and the ability to derive evaluation data, is the nature of the organisational culture that the learner returns to post-course. The Roullier et al. (2006), study showed that culture must be positively affected in order not to defray the benefits of the training. To this end, managers need to be appropriately trained to understand how to support and challenge delegates, both pre and post-course. The findings showed limited training in this area, particularly in

the area of quantification of post course impact, and this is an area for practitioner attention; both to ensure enhanced Transfer as well as Evaluation.

The findings in the lack of success in creating a learner-based culture, were disappointing for both effective transfer and evaluation. The findings suggested that learners left training (etc.) with little, or no, responsibility to implement actions, or to capture that value, despite the fact that their engagement and confidence were improved. This would lead to less effective learning embedding, and negate the value and cost of the development. Bramley and Kitson (1994), stress the need for the learner to have some control over the outcomes from the learning, rather than simply turning up and being expected to 'supply the motivation', and this may be difficult to achieve, given the organisational focused evaluation processes still being operated by the majority of evaluators. The work of Geertshuis et al. (2002), suggested that the role of the learner, in the creation of evaluation outcomes, would be of significant benefit to the change of culture required to make evaluation overall more effective. Anscombe (1959), and Basareb (2007), also stress the need for strong 'intention' from the delegate, to ensure commitment to learning transfer and the need to focus training delivery on the creation of 'practical knowledge' and skills and reduce 'scrap learning' as a key part of this process.

5.4.4 *Reflection on this Phase*

This Phase sought to utilise the views and practice of evaluation practitioners within L&D functions. One of the obvious weaknesses in the research was to fail to include the views of some important stakeholders; although the level of complexity in achieving this would have been challenging. However, the view of appropriate line managers, the most senior HR practitioners, and other external evaluators, would have enriched this section, from the understanding of wider perspectives. Although the focus was this sample, and the sample in itself was strong and legitimate, the findings, in isolating such a group of practitioners,

depended totally on their self awareness and perception of value, and needed the counter balance of whether their opinions were accurate or, even, correct. Research into the perceptions of value of the function within the stakeholder base, would add value for the function, and illustrate the requirements for evaluation.

Another reflection for this section resolved around timing. During this period, the creation of the new framework was underway, and some of the questions may suggest personal bias in attempting to prove a negative, in order to justify the decisions made in the framework. Whilst I was aware of this bias and its subsequent results, it is the role of the reflexive practitioner to debate with themselves whether certain questions had been appropriate, in the light of a more objective research stance. Whilst Gray (2009), maintains that a researcher is not a disinterested observer, and that competence and world view may skew outcomes, the need for pragmatism does appear to challenge a researcher to remember the ethics and methods needed to retain sufficient objectivity in the process.

5.5 Phase Three and Four — Case Study

The Case Study approach was adopted as this phase was predicated on the effectiveness of a new approach to evaluation, based on the ideas of Spitzer et al. (2005), that evaluation could be predictive in approach rather than summative. The new framework removed all aspects of summative evaluation from the process, other than whether specific actions had been achieved, that would lead to outcomes that generated targeted impact.

The creation of the framework built on the ideas of a range of literature sources including: The use of the concepts of the Law of Large Numbers, Bernoulli (1713), in tandem with the work of Cascio and Boudreau (2011), utilising Utility Theory to create a metric 'engine', to understand the measurements, and drive the forecasts. In addition to these sources the Anscombe (1959), ideas around 'intention' were selected to create

the motive for action, that could drive a probability score, to inform the L&D function of the efficacy of the training delivery, and the likelihood of Transfer. In addition to these sources were Hubbard (2014), Walonick (1993), and Seigal (2013) representing the forecasting schools of thought.

The overall research depended on the theory of Hubbard (2007. 2014), on creating a 't-sample', with its own definition of a benchmark (or target) number, against which any claims could be assessed. This removed the need for traditional methods, including Bayesian analysis or regression. The process of the research initially utilised a retrospective sample, as recommended by Armstrong (2001), from which initial forecasts could be generated on data, where a third party's independent result had already demonstrated a benchmark number.

The findings from this 'testing phase' were expected to be within tolerances that had been created from the work of an expert group of evaluators. The process and workings of this group are outlined in Chapter 3. The process of forecasting was built upon the concept that a summative figure would determine the 'actual' ROI number, say 333 per cent. That number became 100 per cent in the forecasting process, against which the forecast would be made. In this no commentary about relative levels of ROI, from different courses or organisations, needed to be considered, simply the forecasted accuracy.

The findings for the expert evaluator group were that a result of 85 per cent, of a target ROI score, would be acceptable, if the reduction in cost and organisational load were sufficiently reduced to justify the lack of accuracy. Also, the forecast should never exceed the benchmark score, to avoid accusations of 'wild forecasting', or a lack of credibility, as recommended by Seigal (2013). The ability to assess cost reduction was not considered in this initial stage, as the focus was more on proving the accuracy of the forecasting process.

A range of scores for the courses forecasted ranged between 86 per cent and 94 per cent against the benchmark of a range of courses, from different organisations which had been previously evaluated using summative data. The lower scores for mandatory training were explained by some poor data input. The coaching prediction was particularly pleasing at a personal level, as it showed that the process could have equal validity for learning interventions other than training. Those initial findings, from the retrospective sample, were deemed satisfactory, and within the tolerances that had been developed from the work of a group of expert evaluators; however, to resolve issues some adjustment of the internal software algorithm well as the accuracy of data input needed were resolved before the actual case study forecasts were carried out, particularly because of the high expectations of the Case Study group.

Despite the work of the advocates of this 'formative only' approach, for example, Spitzer (2004), and Basareb (2007), there was no literature available at the time of this research to be able to create a comparator to these results, hence the creation of a process of performance against expectations.

The actual Case Study forecasted the performance of a wide range of courses and programmes, across a diverse group of contributors, each of whom generated different ROI outcomes from their activities. The forecasting took place in advance of the learning, using the mechanisms developed from both the literature and practice, and the results set aside until the work of the summative evaluation was complete. The Case Study group defined their expectations of a percentage validity score, and this became the internal 't-score' and benchmark, reflecting their individual aspirations and need to be persuaded. Each of the learning interventions was then assessed by an independent third party evaluator to create an accepted 'Gold Standard' benchmark.

The changes in process and accuracy of input were reflected in the accuracy of outputs. Overall accuracy ranged from 87 per cent to 95 per cent. There was a distinct linkage in levels of accuracy and the levels of

current evaluation process that existed within the organisation. For example, where an accuracy of 95 per cent was returned, that organisation had substantial evidence of professional evaluation taking place, and a range of learning embedding processes. This reduced the internal variances and complexity, and allowed the forecasting be to be very accurate, as well as showing the links between good 'transfer' and 'evaluation'.

The Case Study findings showed levels of accuracy acceptable to the group that had determined the target, by operating the forecasting process against the 'Gold Standard' summative process by an independent Third Party evaluator. This is significant within the data, as no results exist for this type of practitioner research within the literature, and represents an approach that could be utilised by practice. However, in order to ensure that it would be possible for the L&D function to use this method, the proof of its accuracy was also supported by significant cost advantages in using this process.

The cost savings showed a huge benefit in using this approach. In many cases the overall cost was around a quarter of the full cost of a professional evaluator, as well as the same levels of reduction in organisational load. Encouragingly, greater competence in using this forecasting approach would certainly further drive down costs, as extra steps were taken in this process because of the nature of the research, and would not normally be required for a commercial application.

The subsequent opinion survey, generated from the views of the sponsors, reflected satisfaction with the result and the cost savings, as well as an encouraging endorsement of the result. Whilst Gray (2009), suggests that legitimate research should be able to replicate this result, this may pose challenges, because of the use of proprietary software to generate the results.

5.5.1 Reflections on this Phase

One of the challenges in creating a new framework is the level of 'proof' needed in order to make a claim of applicability into the wider literature; therefore, Gray (2009), suggests that the areas of Validity and Reliability should be revisited to assist with this process. Validity is helped in this report through the strong links between the literature review and the theoretical ideas and outputs. Both in the valuation and the forecasting literature, the works of Anscombe (1959), Spitzer (1984), Cascio (2001), Seigal (2012), Hubbard (2014), and Gray (2009), are shown as fundamental building blocks in thinking. Associated fields of thought have also served to create fresh perspectives, and have added value in the creation of knowledge and outputs from the Case Study. In addition, 'thick descriptions' have served to outline the context, as well as the outputs from the research. Reliability has been considered by the use of an independent third party evaluator, in order to create an 'internal benchmark', and the use of computer-assisted programmes for coding and analysis. The issue of external reliability is challenging because of the nature and requirements of those that took part in the Case Study, leading to problems in assembling the precise balance of context and data points, however, there is no reason not to apply the framework to different retrospective and fresh data sets, and to follow the same third party process, in order to attempt to replicate any result.

Having established a pragmatic approach at the outset of the research, an early decision was taken to make this stage of the research as pragmatic and rapid as possible. This was driven by a range of needs including: the stakeholders who wanted limited load on their organisation, as well as my personal need to find out the result, and contain the financial impacts of the research, that were being borne by me in paying for a third party evaluator. In addition to this, the 'ticking of the clock', to submit something of value for the Doctoral process, became a driver, to create a containable process that could deliver a defined outcome suitable for the needs of the research.

Whilst this pragmatism drove the approach as well as the statistical justification, with hindsight, more statistical justification could have been

created to satisfy a more 'theoretical audience'. In the practitioner base, 'whether' something works is more important than 'why' it does, and with hindsight, more thought would have been given to the justification of the latter. However, within the practitioner base, the new framework is already in use in a number of organisations, as well as new iterations and variants of the concept. Future researchers are welcomed and encouraged to use the data generated from these forecasting approaches as required.

5.6 Conclusions

In the methods section, I discussed my move away from attempting to create a paradigm into a practical framework, that could become a paradigm if it met the needs of the practitioner base. In the chapter, I posed a number of questions, and it would make sense to discuss those questions, in the light of the learning from both the literature review and the subsequent research; both to check whether the answers have added value, and to ensure that the correct questions were asked in the first place. They were:

Questions	Comments
How can a new evaluation framework built upon Decision Theory principles add value and allow L&D to prove value in a cost effective way?	By creating a competing view of evaluation based on the value of the information it generates. Some of this is clouded by the competing academic views of evaluation whilst recognising the lack of effectiveness within the existing paradigm. The culture of evaluation needs to change as outlined in this report to adding value rather than a cost reduction focus. And finally, the practice needs to move from (not) 'doing Kirkpatrick' to 'doing evaluation'
How can the competing drivers for the need for change and the need for credibility in evaluation be resolved in a function under pressure?	In my view and the view of the small sample of contributors the Decision Theory-driven approach using forecasting worked for them. However, they recognised the challenge in persuading a profession where the leading view is to uphold an idea from 1954 through their professional

	development more highly than other ideas (including their own!)
--	---

A range of sub-questions was refined as follows:

Sub Questions	Comments
What attitudes exist in the practitioner base that drive or restrict attempts to evaluate	A range from time poverty to overwork to Griffin's (2011), worry about internal credibility
Does the outcome required by stakeholder create some of the confusion at the heart of the process	Absolutely yes. But usually only in the minds of the academic base. The practitioner base have to handle the various internal pressures without well researched alternatives to the prevailing paradigm
What part does credibility and resource effectiveness have to play in any solution	A vital part – and this is the heart of the problem. In order to let go of the prevailing paradigm, too big a leap of faith may be required by the practitioner base
Does the longevity of the paradigm represent excellence, apathy or a lack of choice	This research suggests that it represents the absence of a credible alternative. But those other factors exist and there is insufficient pressure on the evaluation thought base to produce new solutions
Should the process of 'what is possible in technology' create the process of evaluation that organisations buy into?	It should not – but the reliance on LMS suppliers may inadvertently be the fulcrum for change – perhaps this is the market to adapt first.
Does the Financial or the Operational community believe that the evaluation of training is credible and/or worth the effort	This research failed to answer this question. Although it must be said that too little pressure comes from outside the L&D function to evaluate and that when the function does come up with new ideas (Utility theory) it is the organisation that often pushes back the findings

5.7 *Suggestions for Practice*

Learning practitioners, on either the supply or demand sides, usually have a very open approach to learning, that seems to be less obvious when investigating the subject of evaluation. One of the problems appears to be that there is no established Journal of Evaluation dedicated to the L&D arena – one exists for Social Impact Projects, but the L&D community needs in this area have become subsumed in HR, a function with arguably even less of a view or competence in their ability to understand their own value.

In addition to this, there needs to be greater development of analytical and judgemental skills within the L&D function. Whilst the CIPD now has this as a small part of the syllabus for new practitioners, we may simply have to wait some time for the results of this to feed through; although I suspect that a wider culture change needs to permeate L&D, with a move back towards reclaiming the tool-kits and approaches of OD, vacated since the launch of the Ulrich framework.

One of the biggest challenges may be the need to change the culture of L&D, so that evaluation is seen as a worthwhile activity, that adds value and insight into the process of HR. In order for this to be achieved, the CIPD should immediately reconsider its support for the prevailing paradigm and decide to propagate new models and ideas of evaluation within the function, in effect, initiate a form of positive discrimination for other evaluation ideas and concepts. This positive reinforcement would give comfort to a functional area, where different approaches have an external seal of approval and legitimacy, against a professional standard. At the moment, I suspect, somewhat pessimistically, that the body, perhaps naturally cautious and risk averse (hence the practitioners it produces), would ally itself with one of the more established organisations in the field, and hence reinforce the prevailing paradigm still further.

One of the simplest ways for people in practice to be affected by ideas is for them to read and experiment with many of the concepts, in this document and, in the publication of many of the key thinkers in the field and to this end I intend to publish a guide for practitioners around this subject in the future. The frameworks from the literature may have to be made more palatable for a non-academic audience but, perhaps through publishing or engaging with some form of eLearning or a software solution, there could grow a group of challenging, enlightened individuals, who can make the change to the paradigm from within the function.

On the supply side of the training industry, including training companies and in-house suppliers (who have much to gain from using this approach), perhaps these organisations should be the target audience for a change in perception and approach, rather than the internal teams. After all, they have a vested interest in proving and showing value, and have the most to gain from demonstrating their credibility. Collaborating with suppliers may well be the best route to test the framework more extensively in the future. Malcolm Gladwell, in the Tipping Point, suggested that a new idea needed an originator, but most importantly a 'maven' to socialise the ideas, and perhaps, the structures of social media, and the third party suppliers could be those mavens, and could help engagement with the concepts.

The area of Learning Transfer still has much traction and needs to be better understood by the L&D function. It seems easier for L&D to access the concepts and its processes, and having fewer metrics could be the accessibility point the evaluation world needs to change the culture. As well as this area, a number of ideas around Big Data and metrics surfaced in the Literature Review, and the rise of computing power to help drive fuzzy logic solutions could help evaluators build even more interesting software. Finally, remembering the associated field of value creation in Marketing is useful, and understanding both online and offline metrics currently being investigated by Marketers could add real value.

One final point is that having encountered the methods and processes involved with coaching, facilitation, meetings management, events and conference organisation, they all tend to replicate the same approaches and (particularly with coaching), are using similar ideas and constructs to help them identify value – perhaps more cross-silo working would help the L&D professional touch base with the worlds of operational, marketing, sales, management, and research.

5.8 Suggestions for Future Research

There are a number of areas that would be interesting ideas for research:

- Further examination and legitimisation of the concepts of formative and stochastic evaluation as built upon in this document
- Link training transfer, with tangible evaluation metrics, to show the outcomes from better transfer, these areas of research are mutually beneficial and have interchangeable application in practice
- Links from Engagement and Productivity into perceived value, to show the causal links between 'soft' training and tangible results
- Investigate the effects of machine learning and fuzzy analytics to transform measurement, and use that learning in the L&D function
- Identify a better metric than ROI – surely the reliance on archaic measures and processes could be challenged
- Investigate the extent to which Decision Theory can generate even more tools and frameworks, or those which could be drawn into evaluation from complementary fields of theory
- Investigate the rise and rise of Virtual Reality and determine how this will either revolutionise, or replace, the L&D function.

All in all, I am more optimistic, as a member of the L&D practitioner base, when I ended this process than when I began, as I believe there are significant gains to be made, both in better learning, and more effective evaluation. It is for the next generation to turn that optimism into tangible and practical approaches and products.

WORD COUNT: 61,991

Chapter 6: Learning Reflection

At the end of any project, good practice dictates there should be a period of reflection and a discussion about the learning journey I had experienced, that can be both useful for me and a point of development for me to bear in mind in practice and further research in the future.

I had three objectives when starting the DProf process;

- I wanted to create a new framework for evaluation, that would quickly add value and 'make a difference', perhaps even replacing the prevailing paradigm
- I wanted to enjoy the process of learning, whilst having the structure and rigour of having to create an output
- I wanted the recognition of a qualification that reflected the size of the journey I had taken and the achievement of the evaluation output

In hindsight, it is clear that these are conflicting objectives. The first assumes a speedy process, and the second a more reflective and deliberate set of processes. The other conflicting factors were the need to build a commercial practice in a difficult trading environment; to devote significant time to the process of study and reflection, as well as the need and desire for personal stimulation, debate, and active learning, being in conflict with the process of distance learning.

In truth the objectives, real life, and the pace of the research itself conspired to break the research journey into discrete segments that included all - work, study, and the resolution of some personal challenges:

- Segment One was based around the original opinion survey process and some reading around the core subject area
- Segment Two was a period where I lost both my parents and the trading climate was in crisis – virtually no studying took place here

- Segment Three followed a period of re-engagement with the studying process and I ran the deep-dive research process
- Segment Four followed a period of serious illness and the divestment of my training business and establishment of a new entrepreneurial activity. This period coincided with the Case Study element of the research

Throughout most of the life of the process, I constantly engaged in reading and, every now and then, came across a definitive idea or text that really shaped my thinking. One of the issues with this 'shaping' was often the resulting diffusion and 'blind alleys' I followed, before I could re-engage with my actual subject. Whilst challenged by a colleague that learning 'really shouldn't be so chaotic', that approach works best for me in order to find unusual links and themes in the underexplored areas of knowledge. I believed, from the beginning, that the answer I was looking for, to create a new paradigm, was in a non-aligned field of knowledge, so wider reading seemed a sensible idea. Perhaps this initial belief was more important than I realised, by enabling me to find greater enlightenment and inspiration in the field of thought outside of the narrow focus of evaluation. As a result the literature review was a painful, ill-disciplined process, until I suddenly found enlightenment from the chaos in the research subject.

The key works in Segment One, as well as the core evaluation thinkers that were known to me, were the ideas of George Kelly and Personal Construct Theory. This led to me spending many hours re-acquainting myself with Repertory Grid – something I had toyed with in the 1980s'. Sadly, whilst I enjoyed the diversion, this added little direct value to my actual primary objective. In addition, engagement with the works of the Human Capital field of study e.g. Fitz-enz (1988 etc.), added some value, but mainly introduced me to my core measurement text by Douglas Hubbard (2014). From this work came my interest in calibrated estimation and t-scores, as well as introducing me to Rasch and his ideas around assessment, although those ideas were totally distracting and of little value.

Segment Two was something of a formal learning desert, other than some texts around the wider subject of learning. However, I continued to absorb learning widely but, because of the context of the time, this was primarily through podcasts. Therefore, I was inspired and influenced by Tim Ferris, Seth Godin, Stephen Dubner, Jeremy Frandsen, et al., as some of the more interesting, and new measurement analytics and ideas began to spill out of internet technologists, entrepreneurs and marketers.

The key works in Segment Three in the evaluation field were the works by Cascio and Boudreau, and the development of Utility Theory. Non-core reading was stimulated by David Rock, and his work around Neuroleadership. A genuinely fascinating subject, following this train of thought, allowed me to engage with some new ideas around learning theory and management practice. In the end, though, whilst building more knowledge in the area of Working Memory, little was transferred into the final field of research. However, the coaching session I had with him, as part of this process, felt like an intellectual pummelling, and worth every moment.

Segment Four was the full re-engagement of the process, stimulated in part by my academic supervisor, as well as the sparking of new ideas - again through the works of Tim Ferris. I became highly influenced by thinkers, including Richard Griffin and GE Anscombe, in beginning to think differently about the whole concept of evaluation. As someone who values innovation and creativity, and has those processes at the heart of my professional practice, it is not surprising to find that this flowering of interest, creation, and written production became my favourite segment.

The work of Anscombe forced me to contemplate, to battle through the text and develop a steely resolve not to be beaten by the style, but ultimately to engage with her ideas at a deeper and more thoughtful level. I have found it difficult to articulate some of the ideas in the work without making them sound trite and simplistic, but my engagement with a form of 'practical philosophy', and her theory adaptation has been one

of the real joys of the whole learning experience. Few books have made such an impact on me as this one!

During the research phase I met another evaluation expert who was completing Masters research into learning embedding and Utility Theory, and this proved to be an interesting seam of activity. My natural curiosity was to dip into this field and, whilst it opened my eyes into another approach within the L&D sphere of practice, it did not appear to offer anything really new at that time, until I was able to link back into the work of Cascio, and see the possibility of an innovative use for a previously 'unloved' concept within organisations.

I have described the process of framework formulation in the Methods' section, but really the process was a significant part of my learning journey, and led me to reflect on the need for learning to be an active process. In my epistemology section, I had identified closely with the ideas of pragmatism and, during the life of the project, this served me well. However, I do think I made trade-offs across the life of the research that did not serve the needs of the research as well as it could or should have done. The pragmatic approach works well when combining with other researchers, or having plenty of support in the learning process, however, the absence of support (other than my academic supervisor and support that I had arranged), was a real issue, and made the process too pragmatic and risked become a process of 'cranking the handle', towards the end of the learning journey in order to deliver the output. In a way, it felt as though I was involved in a pseudo PhD process whilst on a DProf track, and this seems to be a missed opportunity for my own learning.

One of the concepts I did enjoy reinvestigating was the area of 'deep learning', where total immersion can take place, and the time and space is created for focus and concentration. The ideas of Stephen Pressfield helped here and, during the course of writing up, they certainly helped keep me on track and aware of the learning and perspective I was gaining.

One of the key decisions I took early on was to adopt the services of a 'critical friend', whose role was to challenge the research data and the conclusions drawn from it. Her role was also to ensure that any commercial interests did not skew the data collection, or any other conclusions. The person I identified had a PhD completed in a complementary subject area, and was a supervisor for the Open University, and acted as a brake on the many enthusiasms, ideas and diffusions, which I have had as a natural part of my learning style. She did her best to help me manage the internal actor – observer – researcher continuum but, most of the time, I operated on my own, and often found the whole learning processes deeply frustrating and a missed opportunity to share more of the learning, and have more challenge through the process.

One of the challenges of 'distance learning' is the loneliness of the journey and the seeking for answers – one area of enlightenment for me has become the realisation that learning is really the formulation and seeking of questions to aid understanding. One of my huge areas of frustration was the design and creation of the Literature Review – after all, hundreds and thousands are written every year – so why isn't there a template? Having completed the chapter in this document, I realise that was the wrong place to start and the worst question of all to ask and this realisation is, for me, the stepping up from Masters to Doctorate level.

My academic supervisor, provided by the university, has been a source of inspiration, using practical coaching techniques, to point, shove and nag me into producing outputs. I thoroughly enjoyed the interaction, fun, and challenge from our sessions, and deeply regretted the lack of interaction with other academics in the university. I realise that workshops were provided, from time to time, but the issue for entrepreneurs working full time in one of the most challenging business environments in recent times, is the lack of available resources and availability to travel, to take part in traditional workshop-based learning. Where I did attend, the

sessions were useful and enjoyable – I still am regularly distracted by the ideas associated with hermeneutics that came from a specific session.

In using hindsight to aid reflection, perhaps having more structure in the process, would have helped me more. More checking in, and output production, at an earlier stage, would definitely have helped, as only when I was engaged did I really learn – an obvious point, but something that is key when thinking as an evaluator. Of course, I should have created that but, perhaps, also the university could build this into its process.

The sobering thought during the course of the research was the realisation of the inability to actually change or affect the paradigm. In addition to the work of Kuhn on paradigms, the Lindy Effect establishes the idea that something that has a past 'will have an equivalent future'. Mandelbrot (1984) further developed the term from the Goldman (1964) concept, based in the media, to explain the geometry of nature linked to future life expectancy. Taleb (2002) explained that this added to the 'antifragile', and explained why certain ideas, concepts, companies etc. have an inbuilt longevity, where no natural entropy exists. He links the concept to 'survivorship bias', which relates to attributing overall results for the one or two examples of things that are seen, for example, 'music was better in the past', because the few good things are remembered selectively and the vast mass of 'other music' is ignored....

This means that all I can do is what others have done before me – to create my approach, test it and see what the market makes of it. Paradigms create themselves rather than become created, so my goal remains the same, but the strategy for achieving it must change.

As an evaluator, it would be a simple process now to calculate the ROI of the DProf but, perhaps, I should simply reflect on the second objective, and consider this as a means of helping me achieve that. The process of learning cannot be undervalued, and I have gained skills and perspectives, as well as ideas for innovation in business practice, that will

last for a while to come. The practice of 'thinking' is a key realisation and by continuing to build processes and perspective into practice, some of the key learning from this research process can begin to permeate and illuminate my professional practice.

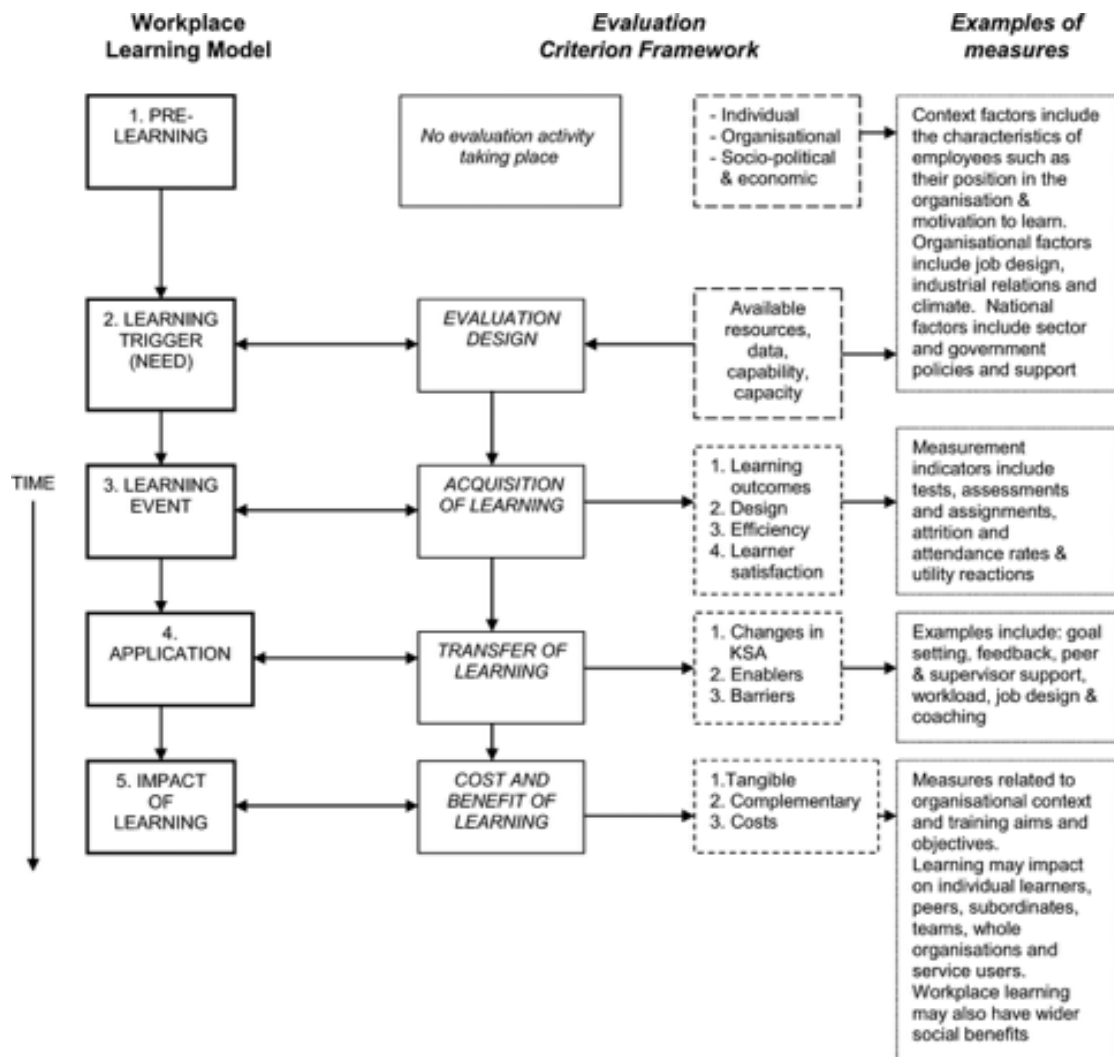
My objectives now are to build on this learning process, and do more – perhaps learning and masochism are part of the same process, but it does seem a shame to have endured the process, and not be able and willing to contribute more in the future, perhaps to even to begin the next Doctorate.

Appendix 1 — Guerçi and Vinante (2011) — Framework

Framework of Stakeholder Needs and Drivers for training evaluation, and the resulting importance for each.

Factor	Relative Importance
The quality and level of knowledge and skills acquired by participants	<i>More important for trainers than for participants.</i>
The number of bureaucratic procedures imposed on participating companies	<i>More important for training providers than for trainers.</i>
The impact of the training program on company results	<i>More important for companies than for trainers, but is important even for participants</i>
The transparency of the mechanism controlling access to financed training services	<i>More important for training providers and companies than for participants</i>
The improvement in the training providers' image among companies	<i>More important for training providers and trainers than for participants.</i>
The possibility to define training financing procedures with the public authorities	<i>More important for training providers than for trainers and participants.</i>

Appendix 2 — Richard Griffin (2011) Evaluation Framework



Appendix 3 — Phase One — Survey Questions

Do you formally evaluate learning in your organisation?

- Yes - No

If yes - to what extent?

Do you use the same approach for all programmes?

- Yes - Mostly Yes - Mostly No - No

What technology do you use when evaluating?

To what extent does your LMS drive the evaluation you do?

If no - why not?

How do you know a programme has been effective?

How would you describe your attitude to evaluation?

- Very Positive – Positive – Neither Positive or Negative – Negative
– Very Negative

Other Comments

Appendix 3a — Phase Two — Contributor Briefing Document

Interviewee Briefing Sheet

Thank you for agreeing to take part in this research process which is part of my Doctoral programme at the University of Middlesex.

The purpose of this document is to brief you of the aims, process and ethical considerations of the research.

By agreeing to schedule and attend the meeting, or complete the research questionnaire, you have tacitly read, understood and agreed to the methods and mode of operation of the process that have been discussed in advance of this note.

Please feel free to contact me in advance of our meeting should you have *any* issues, questions or queries.

What is the aim of the meeting?

The research I am conducting is attempting to create a new framework for the evaluation of L&D programmes. As someone with expertise and experience in either carrying out evaluation, your views are vital in informing the research by collecting your views and opinions as well as the current practice you operate.

This research process aims to collect your views, opinions and experiences through the medium of informal interviewing or questionnaire completion.

What is the Process?

As agreed, our meeting should take us no longer than 90 minutes and will cover the following areas:

Your views on

- the evaluation of training overall?
- The blockers which exist to successful evaluation of training?
- The effectiveness of the evaluation of recent courses?
- Which areas should/should not be evaluated?
- the usefulness of the outcomes in changing on-going behaviour (or the achievement of other objectives)?
- How could evaluation be improved?
- Whether evaluation worth the money?
- The decisions evaluation helps you make?

Other points you would like to make....??

FAQ's

How will my answers be treated in terms of confidentiality?

- You can determine the level of confidentiality depending on your personal comfort and organisational processes, particularly with regard to specific names, projects, company names etc. At this

stage, the proposal is to offer complete confidentiality to all contributors both in terms of personal or organisation name.

- Any names of material which could identify a specific individual will be removed as part of our own checking process
- A transcription of the meeting can be provided to ensure you agree that we have captured a fair and accurate representation of your comments
- All notes and transcripts will be destroyed after 12 months of the meeting taking place
- The notes and outputs will only be used for the purposes of the Doctoral research
- You can have full sight of the completed Doctoral project

How will my answers help you?

The answers will help:

- Create with the development of a new framework
- Build on metrics that actually add value to L&D
- Understand the 'real world' challenges of current practice

Appendix 4 — Phase Two — Survey Questions

What is your attitude to evaluation?

- Positive – Negative - Ambivalent

What types of training do you evaluate?

- All – Core skills – Professional Qualifications & CPD – Personal Development – Organisational initiatives – Mandatory – e-learning - Coaching

Why do you evaluate?

Which evaluation method/s do you use?

What is the data from your evaluation process used for?

What decisions are made based on the data within your reports?

In your view, how are the reports/data regarded by stakeholders?

What process do you operate when a request for training reaches the function?

Do you create an outcome target for a programme?

Do you identify learning points for each programme?

Do you identify post course actions for the delegates for each programme?

I believe that evaluation should be a key factor in the design of a programme

- Strongly Agree – Agree – Neither Agree or Disagree – Disagree – Strongly Disagree

I believe that evaluation should be a key factor in the delivery of a programme

- Strongly Agree – Agree – Neither Agree or Disagree – Disagree – Strongly Disagree

To what level are you interested in confidence and/or engagement as a course output?

- Strongly Interested – Interested – Neither Interested nor Disinterested – Disinterested – Strongly Disinterested

What approaches are used to embed to learning?

What new technologies (e.g. social media etc.) are used to help embed learning?

Indicate areas in which managers have had development?

What percentage of responses do you receive post course?

What methods have you used to stimulate responses (most and least effective)?

What methods have you used to foster a culture where the delegates voluntarily evaluate their own learning?

How is the L&D function evaluated?

How is the budget for development structured?

What new approaches to learning are being implemented within the organisation?

How are the results of those new approaches being evaluated?

How did you learn to evaluate?

Outputs screened by request – What returns do you achieve from the following types of learning?

- All – Core skills – Professional Qualifications & CPD – Personal Development – Organisational initiatives – Mandatory – Elearning - Coaching

Appendix 5 – Phase Three – Expert Workshop Overview

Evaluation Forecasting Method Discussion

Date: Tuesday Jan 6th 2015 – Venue: Kings Place, London

Host: RT: Facilitator – Trevor B

Participants: Jackie, Guy, Nigel, Steve S, Steve P, Valerie, Prof WJ, Dr B, Kate, Tim, Michael, Dr C.

Time	Session Activity	Key Questions to Answer
8.30 – 9.15	Welcome, Introductions, Ground rules, Ethics	How to be sufficiently challenging How to avoid being too supportive or helpful
9.15 – 10.30	RT Presentation – Academic research to date. Introduction of new framework. Q&A	Queries or points of clarification
10.30 – 13.00	Group activities (3 groups of 4)	Does the framework make sense Alternative approaches within the broad concept What evidence supports the framework Blockers and Drivers to adoption Case Study ideas Forecasting guidance and knowledge share Voting Positive/Negative
13.30 – 15.30	Feedback to RT (captured by facilitator)	
15.45 – 16.30	Group work and Feedback	What impact does this feedback generate – guidance for responses to the feedback
16.30 – 17.20	Forecasting validity guidance (4 groups of 3)	Level of accuracy guidance – thoughts and expertise share Case Study question creation
17.20 – 17.30	Final Thoughts Next stage of process Close	

Appendix 6 — Phase Four — Survey Questions

Before the experiment, I believed the new method would work?

What do you think are the benefits of this new approach?

What do you think are the risks of this new approach?

Do you believe that the cost savings are an acceptable trade off for the reduced precision of the framework?

I now trust this approach?

We will carry on with this new approach?

Appendix 7 – List of References and Bibliography

- Abernathy, D. J. (1999) Thinking outside the evaluation box, *Training & Development*, 53(2), 19–23.
- Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. (1985) A learning algorithm for Boltzmann machines, *Cognitive Science*, 9(1), 147–169.
- Ackman, D. (2002) Pay Madness at Enron, *Forbes*, 22 March, <http://www.forbes.com/2002/03/22/0322enronpay.html> [retrieved 16 December 2015].
- Adler, A. S. & Kwon, S. (2002) Social capital: prospect for a new concept, *Academy of Management Review*, 27(1), 17–40.
- Adler, M. & Ziglio, E. (1996) *Gazing into the Oracle: The Delphi Method and its application to Social Policy and Public Health*. London: Jessica Kingsley Publishers.
- Aguinis, H. & Kraiger, K. (2009) Benefits of training and development for individuals and teams, organisations and society, *Annual Review of Psychology*, 60(4), 51–74.
- Airasian, W. P., & Miranda, H. (2002). The Role of Assessment in the Revised Taxonomy. *Theory Into Practice*, 41(4), 249–254.
- Akrofi, S. Clark, N. & Vernon, G. (2011) Validation of a comprehensive executive learning and development measure, *International Journal of Training and Development*, 15(3), 184–209.
- Alkin, M. C., Hofstetter, C. H. & Ai, X. (1998) Stakeholder concepts in program evaluation, in Reynolds, A. & Walberg, H. (eds), *Advances in Educational Productivity*, 7, 87–113. Greenwich, CT: JAI Press.
- Alliger, G. & Janak, E. (1989) Kirkpatrick's levels of training criteria: 30 years later, *Personnel Psychology*, 42, 331–42.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H. & Shotland, A. (1997) A meta-analysis of the relations among training criteria, *Personnel Psychology*, 50(2), 341–58.
- Alshibly, H. H. (2015) Customer Perceived Value in Social Commerce: An Exploration of its Antecedents and Consequences, *Journal of Management Research*, 7(1), 17–37.
- Alvarez, K., Salas, E. & Garofano, C. (2004) An integrated model of training evaluation and effectiveness, *Human Resource Development Review*, 3, 386–416.

- Ames, C. & Archer, J. (1988) Achievement goals in the classroom: Students' learning strategies and motivational processes. *Journal of Applied Psychology*, 76, 478–487.
- Ammons, D. N. & Niedzielski-Eichner, P. A. (1985) Evaluating supervisory training in Local Government: Moving beyond concept to a practical framework, *Public Personnel Management*, 14, 211–230.
- Anderson, J. A., Silverstein, J. W., Ritz, S. A. & Jones, R. S. (1977) Distinctive features, categorical perception, and probability learning: some applications of a neural model, *Psychological Review*, 84, 413–451.
- Anderson, V. (2007) *The Value of Learning — A new model of value and evaluation*. London: CIPD.
- Anderson, V. (2009a) Desperately seeking alignment: Reflections of senior line managers and HRD Executives, *Human Resource Development International*, 12(3), 263–277.
- Andersson, U., Forsgren, M. & Holm, U. (2002) The strategic impact of external networks: subsidiary performance and competence development in the multinational corporation, *Strategic Management Journal*, 23, 979–96.
- Anscombe, G. E. M. (1957) *Intention*, 2nd ed. Cambridge, MA: 2000.
- Australian Public Service Commission. (2005) *Evaluating Learning and Development: A framework for judging success*. Canberra: APSC.
- Aragon-Sanchez, A., Barba-Aragon, A. & Sanz Velle, R. (2003) Effects of training on business results, *International Journal of Human Resource Management*, 14(6), 956–980.
- Arbuckle, J. L. (2003) *Amos 5.0 Update to the Amos User's Guide*. Chicago, IL: SPSS.
- Armstrong, J. S. (1980). *Unintelligible Management Research and Academic Prestige*, http://repository.upenn.edu/marketing_papers/117 [retrieved 14 12 2014].
- Armstrong, J. S. (2001) Evaluating Forecasting Methods, in Armstrong, J. S. (ed) *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Boston, MA: Kluwer Academic.
- Arrow, K. (1973) Social responsibility and economic efficiency. *Public Policy*, 21, 300–317.

- Arthur, W., Bennett, J., Edens, P. S. & Bell, S. T. (2003) Effectiveness of training in organisations: a meta-analysis of design and evaluation features, *Journal of Applied Psychology*, 88(2), 234–245.
- ATD Research. (2013) *2013 State of the Industry Report*. Alexandria, VA: ATD Press.
- Avey, R. D., Maxwell, S. E. & Salas, E. (1992) The relative power of training designs under different cost configurations, *Journal of Applied Psychology*, 77, 155–60.
- Baddeley, A. D. (1999) *Essentials of human memory*. Hove: Psychology Press.
- Bădescu, M. & Loi, M. (2010) *Participation in Training of Adult Workers in European Countries — Evidence from Recent Surveys*. Luxembourg: Publications Office of the European Union.
- Baldwin, T. T. & Ford, J. K. (1988) Transfer of training: a review and directions for future research, *Personnel Psychology*, 41, 63–105.
- Bandura, A. (1977) *Social Learning Theory*. New York, NY: General Learning Press.
- Barrett, F. J. (1995) Creating Appreciative Learning Cultures, *Organisational Dynamics*, 24(2), 36–48.
- Barsky, A. (2007) Understanding the ethical cost of organisational goal-setting: A review and theory development, *Journal of Business Ethics*, 81(1), 63–81.
- Bartov, E., Givoly, D. & Hayn, C. (2002) The rewards to meeting or beating earnings expectations, *Journal of Accounting and Economics*, 33(2), 173–204.
- Baruch, Y. & Hunt, J. W. (2003) Developing top managers: the impact of interpersonal skills training, *Journal of Management Development*, 22(8), 729–52.
- Basareb, D. (2011), *Predictive Evaluation*. Berrett-Koehler Publishers Inc. San Francisco.
- Bassi, L., Benson, G. & Cheney, S. (1996) The top ten trends, *Training and Development Journal*, 50(1), 28–42.
- Bates, R. (2004) A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence, *Evaluation and Program Planning*, 27, 341–7.

- Bates, R. A., Holton, E.F., III, Seyler, D. A. & Carvalho, M. A. (2000) The role of interpersonal factors in the application of computer-based training in an industrial setting, *Human Resource Development International*, 3, 19–43.
- Bazerman, M. H. & Chugh, D. (2006) Decisions without blinders, *Harvard Business Review*, 84(1), 88–97.
- Beach, Dale S. (1980) *Personnel: The Management of People at Work*, 4th edn. New York, NY: Macmillan Publishing Company.
- Beattie, R. S. (2006) Line managers and workplace learning: Learning from the voluntary sector, *Human Resource Development International*, 9(1), 99–119.
- Becker, B. E., Huselid, M. A. & Ulrich, D. (2001) *The HR Scorecard: Linking People, Strategy and Performance*. Boston, MA: Harvard Business School Press.
- Becker, G. S. (1975) *Human Capital Theory*, 2nd ed. New York, NY: Columbia University Press.
- Becker, L. J. (1978) Joint effect of feedback and goal setting on performance: A field study of residential energy conservation, *Journal of Applied Psychology*, 63(4), 428–433.
- Beckett, D. & Hagar, P. (2002) *Life, Work and Learning: Practice in Post Modernity*. London: Routledge.
- Beeson, J. (2010) *The Unwritten Rules: The Six Skills You Need to Get Promoted to the Executive Level*. San Francisco, CA: Wiley.
- Beharrell, P. (1993) AIDS and the British Press, in Eldridge, J. (ed.), *Getting the Message: News, Truth and Power*. London: Routledge.
- Bennett, N. (1997) The voices of evaluation, *The Journal of Continuing Education in the Health Professions*, 17, 198–206.
- Bentler, P. M. (1990) Comparative fit indexes in structural models, *Psychological Bulletin*, 107, 238–246.
- Bentler, P. M. & Bonett, D. G. (1980) Significance tests and goodness of fit in the analysis of covariance structures, *Psychological Bulletin*, 88, 588–606.
- Bernstein, Allen L. (1966) *A handbook of statistical solutions for the behavioral sciences*. New York, NY: Holt, Rinehart and Winston.
- Billett, S. (2007) Constituting the workplace curriculum, *Journal of Curriculum Studies*, 38(1), 31–48.

- Billett, S. & Sommerville, M. (2004) Transformation at work: identity and learning, *Studies in Continuing Education*, 26(2), 309–326.
- Birnbrauer, H. (1987) Troubleshooting your Training Programme, *Training and Development Journal*, 41(7), 53–61.
- Blanchard, P. N. & Thacker, J. (2007) *Effective Training: Systems, Strategies and Practices*. Upper Saddle River, NJ: Pearson Education.
- Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. New York: David McKay Company.
- Blume, B. D., Ford, J. K., Baldwin, T. T. & Huang, J. L. (2010) Transfer of Training: A meta-analytic review, *Journal of Management*, 36, 1065–1105.
- Blumer, H. (1969) *Symbolic Interactionism; Perspective and Method*. Englewood Cliffs, NJ: Prentice-Hall.
- Boaden, R. J. (2006) Leadership development: does it make a difference?, *Leadership & Organisation Development Journal*, 27(1), 5–27.
- Boeker, W. (1997) Executive migration and strategic change: the effect of top manager movement on product-market entry, *Administrative Science Quarterly*, 42, 213–36.
- Bold, E. (2011) Instruments and techniques used in the design and implementation of change management, *Journal of Advanced Research in Management*, 1(3), 5–13.
- Botchkarev, A. & Andru, P. (2011) A Return on Investment as a metric for evaluating Information Systems: Taxonomy and Application, *Interdisciplinary Journal of Information, Knowledge, and Management*, 6.
- Boudreau, J. (1983) Economic considerations in estimating the utility of human resource productivity improvement programmes, *Personnel Psychology*, 36, 551–576.
- Boudreau, J. (1984) Decisions theory contributions to HRM theory and practice, *Industrial Relations*, 23, 198–217.
- Boudreau, J. & Berger, C. J. (1985) Decision theoretic utility analysis applied to employee separations and acquisitions, *Journal of Applied Psychology*, 70, 581–612.

- Boudreau, J. (1996), The Motivational Impact of Utility Analysis and HR Measurement, *Journal of Human Resource Costing & Accounting*, 1(2), 73–84.
- Boudreau, J. W. & Ramstad, P. M. (1997) Measuring Intellectual Capital: Learning from Financial History, *Human Resource Management*, 36, 343–356.
- Bowles, S., Cunningham, C. J. L., Gabriel, M., De La Rosa, G. M. & Picano, J. (2007) Coaching leaders in middle and executive management: goals, performance, buy-in, *Leadership & Organisation Development Journal*, 28(5) 388–408.
- Brache, A. P. & Rummler, G. A. (1997) Managing an Organisation as a System, *Training*, 34(2), 68–74.
- Bramley, P. & Kitson, B. (1994) Evaluating training against business criteria, *Journal of European Industrial Training*, 1, 10–14.
- Brandon, P. R. (1998) Stakeholder participation for the purpose of helping ensure evaluation validity: bridging the gap between collaborative and non-collaborative evaluations, *American Journal of Evaluation*, 19, 325–337.
- Brandon, P. R. (1999) Involving program stakeholders in reviews of evaluators' recommendations for program revisions' *Evaluation and Program Planning*, 22, 363–372.
- Brews, P. J. & Hunt, M. R. (1999) Learning to plan and planning to learn: resolving the planning school/learning school debate, *Strategic Management Journal*, 20, 889–913.
- Brinkerhoff, R.O. (1988) An integrated evaluation model for HRD, *Training & Development Journal*, 42(2), 66–68.
- Brinkerhoff, R. O. (1989) *Evaluation Training Programs in Business and Industry*. San Francisco, CA: Jossey-Bass.
- Brinkerhoff, R.O. (2003) *The Success Case Method*. San Francisco, CA: Berrett-Koehler.
- Brinkerhoff, R. O. (2006) Increasing impact of training investments: An evaluation strategy for building organisational capability, *Industrial and Commercial Training*, 38(6), 302–307.
- Broadbridge, A. (2004) It's not what you know, it's who you know, *Journal of Management Development*, 23(6), 551–562.
- Brogden, H. E. (1949) When Testing Pays Off. *Personnel Psychology*, 2, 171–185.

- Brogden, H. E. & Taylor, E. K. (1950) The dollar criterion: applying cost accounting concepts to criterion selection, *Personnel Psychology*, 3, 133–154.
- Brookfield, S. D. (1995) *Becoming a Critically Reflective Teacher*. San Francisco, CA: Jossey-Bass.
- Brown, B. L. (2001) Return on investment in training, *Myths and Realities*, No. 16, <http://www.calpro-online.org/eric/docs/mr00031.pdf> [retrieved on 12 Jan 2013].
- Brown, D. C. (1994) *How managers and training professionals attribute causality for results: implications for training evaluation* (Unpublished doctoral dissertation). Urbana-Champaign, IL: College of Education, University of Illinois.
- Brown, K. G. (2005) An examination of the structure and nomological network of trainee reactions: a closer look at 'smile sheets', *Journal of Applied Psychology*, 90(5), 991–1001.
- Brunner, I. & Guzman, A. (1989) Participatory Evaluation: A Tool to Assess Projects and Empower People, *New Directions in Program Evaluation*, 42, 9–17.
- Bryk, A. S. (1983) Stakeholder-based Evaluation, *New Directions in Program Evaluation*, San Francisco, CA: Jossey Bass.
- Bryman, A. (2007) *Quantity and Quality in Social Research*. London: Taylor and Francis.
- Bryman, A. (2008) *Social Research Methods*, 3rd edn, Oxford: OUP.
- Buckley, R. & Caple, J. (1990) *The Theory and Practice of Training*. London: Kogan Page.
- Bumpass, S. (1990) 'Measuring Participant Performance — An Alternative', *Australian Journal of Educational Technology*, 6(2), pp. 99–107.
- Burke-Johnson, R. (1991) Toward a theoretical model of evaluation utilisation, *Evaluation and Program Planning*, 21, 93–110.
- Burke, L. & Baldwin, T. (1999) Workforce Training Transfer: A Study of the Effect of Relapse Prevention Training and Transfer Climate, *Human Resource Management*, 38(3), 227–242.
- Burke, L. & Hutchins, H. (2007) Training transfer, *Human Resource Development Review*, 6(3), 199–206.

- Burke, Michael J. & Frederick, James T. (1986) A comparison of economic utility estimates for alternatives *SDy* estimation procedures, *Journal of Applied Psychology*, 71, 334–339.
- Burkett, H. (2005a) ROI on a shoestring: evaluation strategies for resource-constrained environments: Measuring more with less (Part I), *Industrial and Commercial Training*, 37(1), 10–17.
- Burkett, H. (2005b) ROI on a shoestring: evaluation strategies for resource-constrained environments: Measuring more with less (Part II), *Industrial and Commercial Training*, 37(2), 97–105.
- Burkett, H. (2006) *Ten Tactics for Highly Effective Evaluators*, <http://evaluationworks.net/documents>
- Burt, R. S. & Ronchi, D. (2007) Teaching executives to see social capital: results from a field experiment, *Social Science Research*, 36, 1156–1183.
- Bushnell, D. (1990) Input, process, output: a model for evaluating training, *Training and Development Journal*, 44(3), 41–43.
- Butler, R. (1993) Effects of task-and ego-achievement goals on information seeking during task engagement, *Journal of Personality and Social Psychology* 65, 18–31.
- Button, S. B., Mathieu, J. E. & Zajac, D. M. (1996) Goal orientation in organisational research: A conceptual and empirical foundation, *Organizational Behavior and Human Decision Processes* 67(1) 26–48.
- Byron, J. & Vitalo, R. L. (1991) Quality improvement through exemplar-based productivity analysis, *American Productivity & Quality Center*, Brief 82.
- Cabinet Office. (2003) *Guidance Notes for Policy Evaluation and Analysis*. London: Cabinet Office.
- Cabrera, L. F. and Cabrera, A. (1999) Rethinking utility analysis: a strategic focus, *Business Economics*, Series 13.
- Camerer, C., Babcock, L., Loewenstein, G. & Thaler, R. (1997) Labor Supply of New York City Cabdrivers: One Day at a Time, *The Quarterly Journal of Economics*, 112(2), 407–441.
- Campion, M. A. & Campion, J. E. (1987) Evaluation of an interview skills training program in a natural field setting, *Personnel Psychology*, 40, 675–691.
- Canning, C. (1991) What Teachers Say About Reflection, *Educational Leadership*, 49(3) 18–21.

- Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I. & Mathieu, J. E. (1995) Toward theoretically based principles of training effectiveness: a model and initial empirical investigation, *Military Psychology*, 7, 141–164.
- Carliner, S. (2009) Maybe ROI really is a waste of time, *Training*, 46.
- Carpenter, J., Dickinson, H. (2016) *Interprofessional Education and Training*. Bristol: Bristol University Policy Press.
- Carson, K. P., Becker, J. S. & Henderson, J. A. (1998) Is Utility Really Futile? A Failure to Replicate and an Extension, *Journal of Applied Psychology*, 83, 84–96.
- Cascio, W. F. & Ramos, R. A. (1986) Development and Application of a New Method for Assessing Job Performance in Behavioral Economic Terms, *Journal of Applied Psychology*, 71, 20–28.
- Cascio, W. (1982) *Costing Human Resources*. Boston, MA: Kent Publishing Company.
- Cascio, W. (1982) *Applied Psychology in personnel management*, 2nd edn. Reston, VA: Reston Publishing Company.
- Cascio, W. F. & Boudreau, J. (2011) *Investing in people: financial impact of human resource initiatives*. Upper Saddle River, NJ: Pearson Education, Inc.
- Casey, M. (2006) *Problem-based inquiry: an experiential approach to training evaluation* (Unpublished doctoral dissertation). Akron, OH: University of Akron.
- Cassidy, M., Gory, H. & Stobl, E. (2005) Knowledge accumulation and productivity: evidence from plant level data for Ireland, *Scottish Journal of Political Economy*, 52(3), 344–359.
- Castanias, R. P. & Helfat, C. E. (2001) The managerial rents model: theory and empirical analysis, *Journal of Management*, 27(6), 661–678.
- Cervone, D., Jiwani, N. & Wood, R. (1991) Goal setting and the differential influence of self-regulatory processes on complex decision-making performance, *Journal of Personality and Social Psychology*, 61, 257–266.
- Charan, R., Drotter, S. & Noel, J. (2011) *The Leadership Pipeline: How to Build the Leadership Powered Company*, 2nd edn. San Francisco, CA: Wiley.

- Chartered Institute of Personnel and Development. (2010) *Evaluating Learning and Talent Development*, www.cipd.co.uk/hr-resources/factsheets/evaluating-learning-talent-development.aspx, http://www.cipd.co.uk/binaries/5215_Learning_talent_development_survey_report.pdf, [14 Mar 2012].
- Chartered Institute of Personnel and Development. (2013) *L&D Survey*. London: CIPD.
- Chartered Institute of Personnel and Development. (2014) *L&D Survey*. London: CIPD.
- Chartered Institute of Personnel and Development. (2015) *Training Industry Trends*. London: CIPD.
- Chawla, S. (1995) Conclusion: Reflections of Learning from a Gathering, in Chawla, S. & Renesch, J. (eds), *Learning Organisations: Developing Cultures for Tomorrow's Workplace*, 501–508. Portland, OR: Productivity Press.
- Chen C. T. (1970) *Introduction to Linear Systems Theory*. New York: Holt, Reinhart and Winston.
- Chen, H. T. & Rossi, P. H. (eds). (1992) *Using Theory to Improve Program and Policy Evaluations*. Westport, CT: Greenwood Press.
- Chen, T. Y., Chang P. L. & Yeh, C. W. (2004) A Study of Career Needs, Career Development Programs, Job Satisfaction and the Turnover Intensity of R & D Personnel, *Career Development International*, 9(4), 424–437.
- Cheng, M., Subramanyam, K. R. & Zhang, Y. (2005) *Earnings Guidance and Managerial Myopia* (Working Paper). Los Angeles, CA: University of Southern California.
- Cheng. E. W. L. & Hampson, I. (2008) Transfer of Training: A review and new insights, *International Journal of Management Reviews*, 10(4), 327–341.
- Chiaburu, D. & Lindsay, D. (2008) Can do or will do? The importance of self-efficacy and instrumentality for training transfer, *Human Resource Development International*, 11(2), 199–206.
- Chiva, R., Alegre, J. & Lapiedra, R. (2007) Measuring organisational learning capability among the workforce, *International Journal of Manpower*, 28, (3/4), 224–242.

- Chmielewski, T. & Phillips, J. (2002) Measuring return-on-investment in government: issues and procedures, *Personnel Management*, 31(2), 237–255.
- Clarke, N. (2005) Workplace learning environment and its relationship with learning outcomes in healthcare organisations, *Human Resource Development International*, 8(2), 185–205.
- Clegg, W. H. (1987) Management training evaluations: An update, *Training and Development Journal*, 65–71.
- Clough, P., Earle, K. and Strycharczyk, D. (2009) in Passmore, J. (Ed.), *Psychometrics in Coaching*, Kogan Page, London. VOL. 44 NO. 6 2012 jINDUSTRIAL AND COMMERCIAL TRAINING j PAGE 323
- Coase, R. (1937) The Nature of the Firm, *Economica*, 4(16), 386–405.
- Coghlan, D. and Brannick, T. (2004) *Doing Action Research in Your Own Organisation*. London: Sage.
- Collins, D. B. (2002) Performance-level evaluation methods used in management development studies from 1986 to 2000, *Human Resource Development Review*, 1(1), 91–110.
- Colquitt, J. A., LePine, J. A. & Noe, R. A. (2000) Toward an Integrative Theory of Training Motivation: A Meta-Analytic Path Analysis of 20 Years of Research, *Journal of Applied Psychology*, 85(5), 678–707.
- Cook, T. D., Leviton, L. C. & Shadish, W. R. (1985) Program evaluation, in Lindzey, G. & Aronson, E. (eds), *Handbook of Social Psychology*, 3rd edn, 699–777. New York, NY: Random House.
- Costa, A. & Kallick, B. (1993) Through the Lens of a Critical Friend, *Educational Leadership*, 51(2), 49–51.
- Costello, A. B. & Osborne, J. W. (2005) Best practices in exploratory factor analysis: four recommendations for getting the most from the analysis, practical assessment, *Research & Evaluation*, 10(7), 1–9.
- Coster, P., (1987) The Civil Service Senior Management Development Programme, *Management Education and Development*, 18(3), 204–222.
- Cotton, J. (1995) *The theory of learning strategies: An introduction*. London: Kogan Page.
- Cousins, B. & Whitmore, E. (1998) Framing participatory evaluation, *New Directions for Evaluation*, 80, 5–23.

- Cousins, J. B. & Earl, L.M. (1995) The Case for Participatory Evaluation: Theory, Research, Practice, in Cousins, J. B. and Earl, L. M. (eds) *Participatory Evaluation in Education*, 5–18. London: Falmer.
- Cousins, M. & Stark, J. W. C. (2003) A Predictive Modeling Primer: *Society of Actuaries Annual Meeting*, October, Orlando, FL, Society of Actuaries, New York, NY.
- Coutinho, S. A. (2007) The relationship between goals, metacognition, and academic success, *Educate*, 7(1), 39–47.
- Creswell, J. W. (2002) *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. Thousand Oaks, CA: Sage Publications.
- Creswell, J. W. (2007) *Qualitative inquiry and research design: choosing among five approaches*, 2nd ed. Thousand Oaks, CA: Sage Publications.
- Cronbach, L. J. & Gleser, G. C. (1965) *Psychological tests and personnel decisions*, 2nd edn. Urbana, IL: University of Illinois.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D.F. & Weiner, S. S. (1982) *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey Bass.
- Cronshaw, S. F. (1997) Lo the Stimulus Speaks: The Insider's View on Whyte and Latham's "The Futility of Utility Analysis", *Personnel Psychology*, 50, 611–615.
- Cropanzano, R., Byrne, Z. S., Bobocel, D. R. & Rupp, D. E. (2001) Moral Virtues, Fairness Heuristics, Social Entities, and Other Denizens of Organizational Justice, *Journal of Vocational Behavior*, 58(2), 164–209.
- Crotty, M. (1998). 'Research Ethics and data quality: The implications of informed consent' *International Journal of Social Research Methodology*, 9(2): 83-95
- Cullen, J., O'Connor, M. & Mangan, J. (2004) Matching management tools and techniques with management challenges, *Strategy & Leadership*, 32(3), 27–30.
- Cumming, B., et al. (2002) Predictive Modeling (Health), *Record*, 28(2): *Society of Actuaries Spring Meeting*, 24–26 June 2002, San Francisco, CA, Society of Actuaries, New York, NY.
- Cunningham, I. (1986) *Developing Chief Executives: An Evaluation*. Hertfordshire: Ashridge Management College.

- Dahiya, S. & Jha, A. (2011) Review of Training Evaluation, *International Journal of Computer Science and Communication*, 2(1), 11–16.
- Damasio, A. (2000) *The feeling of what happens*. London: Heinemann.
- Damasio, A. R. (1994) *Descartes' Error*. New York, NY: G. P. Putnam's Sons.
- Damasio, A. R. (2003) *Looking for Spinoza*. London: Heinemann.
- Darby, J. A. (2006) Evaluating training courses: an exercise in social desirability?, *Journal of European Industrial Training*, 30(3), 227–239.
- Davidove, A. E. & Schroeder, P. A. (1992) Demonstrating ROI of training, *Training & Development Journal*, 46(8), 70–71.
- Davidson, D. (1980) *Actions, Reasons and Causes, Essays on Actions and Events*. New York, NY: OUP.
- Dawson, R. (1995) Fill this in before you go: under-utilised evaluation sheets, *Journal of European Industrial Training*, 19(2), 3–7.
- Dayal, Ishwar (2001) Measuring Training Effectiveness, *Indian Journal of Industrial Relations*, 339–344.
- de Bono, E. (1970) *Lateral Thinking: Creativity Step by Step*. New York, NY: Harper & Row.
- de Bono, E. (1984) *The Six Thinking Hats*. New York, NY: Harper & Row.
- Deci, E. L. (1971) Effects of externally mediated rewards on intrinsic motivation, *Journal of Personality and Social Psychology*, 18(1), 105–115.
- Deci, E. L. (1975) *Intrinsic motivation*. New York, NY: Plenum.
- Deci, E. L., Koestner, R. & Ryan, R. M. (1999) A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation, *Psychological* 125(6), 627–668.
- Delery, J. E. & Doty, D. H. (1996) Modes of Theorising in Strategic Human Resource Management: Tests of Universalistic, Contingency, and Configurational Performance Predictions, *Academy of Management Journal*, 39, 802–835.
- Deming, W. E. (1998), Foreword, in Moen, R., Nolan, T. and Provost, L., *Quality Improvement Through Planned Experimentation*. New York, NY: McGraw-Hill Professional.
- Department for Business, Innovation and Skills. (2010) *Skills for Sustainable Growth*. London: The Stationery Office.

- Dessinger, J. C. & Moseley, J. L. (2006) The full scoop on full-scope evaluation, in Pershing, J. A. (ed), *Handbook of Human Performance Technology: Principles, Practices, Potential*, 312–330. San Francisco, CA: Pfeiffer.
- Dey, I. (1993) *Qualitative Data Analysis*. London: Routledge.
- DiPiazza, S. A. & Eccles, R. G. (2002) *Building Public Trust: The Future of Corporate Reporting*. Hoboken, NJ: Wiley.
- Dixon, N. (1990) *Evaluation: A Tool for HRD Quality*. San Diego, CA: University Associates.
- Donaldson, T. & Preston, L. E. (1995) The stakeholder theory of the corporation: concepts, evidence, and implications, *Academy of Management Review*, 20(1), 65–91.
- Donovan, P., Hannigan, K. & Crowe, D. (2001) The learning transfer system approach to estimating the benefits of training: Empirical evidence, *Journal of European Industrial Training*, 25(2), 221–228.
- Driscoll, M. & Preskill, H. (1996) The Journey Toward Becoming a Learning Organisation: Are We Almost There?, in Watkins, K. E. & Marsick, V. J. (eds), *Creating the Learning Organisation*, Vol. 1, 67–80. Alexandria, VA: American Society for Training and Development.
- Driskell, James E. (2011) Effectiveness of Deception Detection Training: A Meta-Analysis, *Psychology, Crime & Law*, 10, 1–19.
- DuPlessis, E. (2005) *The Advertised Mind*, London: Kogan Page.
- Dublin, M. (1989) *Futurehype: The Tyranny of Prophecy*. New York, NY: Plume.
- Dunning, D., Johnson, K., Ehrlinger, J. & Kruger, J. (2003) Why people fail to recognise their own incompetence, *Current Directions in Psychological Science*, 12(3), 83–87.
- Dweck, C. S. & Leggett, E. S. (1988) A social-cognitive approach to motivation and personality, *Psychological Review*, 95, 256–273.
- Dyer, L. (1984) Linking Human Resource and Business Strategies, *Human Resource Planning*, 7, 79–84.
- Dyer, S. (1994) Kirkpatrick's Mirror, *Journal of European and Industrial Training*, 18(5), 31–32.
- Easterby-Smith, M. P. V. (1986) *Evaluation of Management Education, Training and Development*. Aldershot: Gower.

- Ebbinghaus, H (1885). *Memory: A Contribution to Experimental Psychology* Originally published in New York by Teachers College, Columbia University. Translated by Henry A. Ruger & Clara E. Bussenius (1913).
- Elliot, A. J. & Harackiewicz, J. M. (1996) Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis, *Journal of Personality and Social Psychology*, 70(3), 461–475.
- Elliot, A. J. & Church, M. A. (1997) A hierarchical model of approach and avoidance achievement motivation, *Journal of Personality and Social Psychology* 72, 218–232.
- Elliott, C. K. & Knibbs, J. R. (1981) The Training Triangle: Expectations of Course Members, Tutors and Sponsoring Organisations, *Journal of European Industrial Training*, 6(1), 13–17.
- Elliott, M., Dawson, R. & Edwards, J. (2009) Providing demonstrable organisational learning and training, *Journal of European Industrial Training*, 33(7), 657–670.
- Ellis, K. (2005) What's the ROI on ROI? *Training*, 42(1), 16–21.
- Erikson, K. T. (1967) A Comment on Disguised Observation in Sociology, *Social Problems*, 14, 366–373.
- Eseryel, D. (2002) Approaches to evaluation of training: Theory and Practice, *Educational Technology and Society*, 5(2), 93–98.
- Everson, H. T. & Tobias, S. (1998) The ability to estimate knowledge and performance in college: A metacognitive analysis, *Instructional Science*, 26, 65–79.
- Eysenck, M. W. & Keane, M. T. (2000) *Cognitive Psychology*, 4th edn. Hove: Psychology Press Ltd.
- Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T. & Kudisch, J. D. (2012) The influence of General Perceptions of the Training Environment on Pretraining Motivation and Perceived Training Transfer *Journal of Management*, 21(1), 1–25?
- Faragher, J. (2014) 70:20:10 — A model approach for learning?, *Personnel Today*, February, <http://www.personneltoday.com/hr/702010-a-model-approach-for-learning/> [retrieved on 18 June 2015].
- Federico, G. & King, G. (2007) *Understanding the Lee-Carter Mortality Forecasting Method*. Cambridge, MA: Harvard University (Center for Basic Research in the Social Sciences).

- Felli, L. & Harris, C. (1996) Learning, wage dynamics and firm-specific human capital, *Journal of Political Economy*, 104(4), 838–868.
- Felstead, A., Green, F. & Mayhew, K. (1997) *Getting the measure of training a report on training statistics in Britain*. Leeds: Centre for Industrial Policy and Performance.
- Felstead, A., Green, F. & Jewson, N. (2011) *The Impact of the 2008–2009 Recession on the Extent, Form and Patterns of Training at Work*. London: The Centre for Learning and Life Chances in Knowledge, Economics and Societies.
- Fenton, N. Bryman, A. Deacon, D. (1998) *Mediating Social Science*. London: Sage.
- Ferketich, S. (1991) Focus on psychometrics: aspects of item analysis, *Research in Nursing & Health*, 14, 165–168.
- Ferris, T. (2007) *The Four House Workweek: Escape the 9–5, Live Anywhere and Join the New Rich*. New York, NY: Crown Publishers.
- Festinger, L. (1957) *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Fetterman, D. (1994) Empowerment Evaluation, *Evaluation Practice* 15(1), 1–15.
- Fetterman, D. (1996) *Empowerment Evaluation: Knowledge and Tools for Self Assessment and Accountability*. Thousand Oaks, CA: Sage.
- Fillery-Travis, A. J. (2014) The framework of a generic DProf programme – a reflection on its design, the relational dimension for candidates and advisers and the potential for knowledge co-creation, *Studies in Higher Education*, 39(4), 608–620.
- Fischer, R. (2011) 'Cross-Cultural Training Effects on Cultural Essentialism Beliefs and Cultural Intelligence, *International Journal of Intercultural Relations*, 35(6), 767–775.
- Fitz-enz, J. (1988) Proving the value of training, *Personnel*, March, 17–23.
- Fitz-Enz, J. (1994). Yes, you can weigh trainers' value. *Training*, 31:54–58, 1994.
- Fitz-enz, J. (2001), *The ROI of Human Capital*, American Management Association, NY, USA.

- Flamholtz, E. (1969) *The Theory of the Measurement of Individuals in an Organization* (Doctoral dissertation). Ann Arbor, MI: University of Michigan.
- Flamholtz E. G., Bullen, M. L. & Hua, W. (2003) Measuring the ROI of Management Development: An Application of the Stochastic Rewards Valuation Model, *Journal of Human Resource Costing & Accounting*, 7(1), 21–40.
- Flavell, J. H. (1979) Metacognition and cognitive monitoring: A new area of psychological inquiry, *American Psychologist*, 34, 906–911.
- Fletcher, J. (1999) *Situation Ethics*. London: SCM Press.
- Flick, U. (2006) *An Introduction to Qualitative Research*, 3rd edition. London: Sage.
- Flippo, B. Edwin (1971) *Principles of Management*. New Delhi: McGraw Hill Company.
- Flynn, D. J. (1992) *Information Systems Requirements: Determination and Analysis*. London: McGraw-Hill.
- Ford, A., Hornsby, J. & Stoutland, F. (eds). (2001) *Essays on Anscombe's Intention*. Cambridge, MA: Harvard University Press.
- Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M. & Salas, E. (1998) Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer, *Journal of Applied Psychology*, 83, 218–233.
- Ford, J. K. & Kraiger, K. (1995) The application of cognitive constructs and principles to the instructional systems design model of training: implications for needs assessment, design, and transfer, *International Review of Industrial and Organisational Psychology*, 10, 1–48.
- Ford, J. K., Quinones, M., Sego, D. & Sorra, J. (1992) Factors affecting the opportunity to use trained skills on the job, *Personnel Psychology*, 45, 511–527.
- Frechtling, J. A. (2007) *Logic Modelling Methods in Program Evaluation*. San Francisco, CA: Jossey-Bass.
- Friedman, A. L. & Miles, S. (2006) *Stakeholder: Theory and Practice*. Oxford: Oxford University Press.
- Friedman, M. (1966) The methodology of positive economics, in *Essays in Positive Economics*. Chicago, IL: University of Chicago Press.

- Fuller-Love, N. (2006) Management development in small firms, *International Journal of Management Reviews*, 8, 175–90.
- Furió, D., González-Gancedo, S., Juan, M.-C., Seguí, I. & Rando, N. (2013) Evaluation of learning outcomes using an educational iPhone game vs. traditional game, *Computers & Education*, 64, 1–23.
- Galinsky, A. D., Mussweiler, T. & Medvec, V. H. (2002) Disconnecting outcomes and evaluations: The role of negotiator focus, *Journal of Personality and Social Psychology*, 83(5), 1131–1140.
- Garavan, T. N. (1995) HRD stakeholders: their philosophies, values, expectations and evaluation criteria, *Journal of Evaluation Industrial Training*, 19(10), 17–30.
- Garaway, G. B. (1995) Participatory evaluation, *Studies in Educational Evaluation*, 21(1), 85–102.
- Gardner, H. (1983) *Frames of Mind*. New York, NY: Basic Books.
- Garland, H. (1983) Influence of ability, assigned goals, and normative information on personal goals and performance: A challenge to the goal attainability assumption, *Journal of Applied Psychology*, 68(1), 20–30.
- Geber, B. (1995) Does your training make a difference? Prove it!, *Training*, 3, 27–34.
- Geertshuis, S., Holmes, M., Geertshuis, H., Clancy, D. & Bristol, A. (2002) Evaluation of workplace learning, *Journal of Workplace Learning*, 14(1), 11–18.
- Gershon, P. (2004) *Releasing resources to the front line: Independent Review of Public Sector Efficiency*. Norwich: HMSO.
- Giangreco, A., Caugati, A., Sebastianao, A. & Bella, D. (2010) Trainees' reactions to training, *The International Journal of Human Resource Management*, 21(13), 2468–2487.
- Giangreco, A., Sebastianao, A. & Peccei, R. (2009) Trainees reactions to training: An analysis of the factors affecting overall satisfaction with training, *International Journal of Human Resource Management*, 20(1), 96–111.
- Gibson, M. (1983) Evaluation Management Development Programmes, *Management Education and Development*, 14(3), 153–169.
- Gilliland, S. W. & Landis, R. S. (1992) Quality and quantity goals in a complex decision task: Strategies and outcomes, *Journal of Applied Psychology*, 77(5), 672–681.

- Glaser, B. G. (1992) *Basics of Grounded Theory Analysis*. Mill Valley, CA: Sociology Press.
- Goddard, R. W. (1990) The Rise of the New Organisation, *MW*, 1 (1), 3–5.
- Goldberg, E. (2009) *The Executive Brain: Frontal Lobes in a Complex World*. Oxford: Oxford University Press.
- Goldman, A. (1964) Television: Lindy's Law, *New Republic*, 150(24), 34.
- Goldsmith, M., Baldoni, J. & McArthur, S. (2010) *AMA Handbook of Leadership*. Broadway, NY: American Management Association.
- Goldsmith, R. F. & Sarno, A. (2009) 10 ROI Pitfalls: *GoPro Management Conference*, DD Mm, Location, Organisation, San Jose, California
- Goldstein, I. L. & Gilliam, P. (1990) Training system issues in the year 2000, *American Psychologist*, 45, 134–143.
- Goldwasser, D. (2001) Beyond ROI, *Training*. 38(1), 82.
- Gomez-Mejia, L. R. & Balkin, D. B. (1992) *Compensation, Organizational Strategy, and Firm Performance*. Cincinnati, OH: South-Western.
- Goode, E. (1996) The Ethics of Deception in Social Research: A Case Study, *Qualitative Sociology*, 19, 11–33.
- Goodrich, T. J. (1978) Strategies for Dealing with the Issue of Subjectivity in Evaluation, *Evaluation Quarterly*, 2(4), 631–645.
- Gordon, W. (2006) What Do Consumers Do Emotionally with Advertising? *Journal of Advertising Research*, 46(1), 2–10.
- Gray, D. E. (2009) *Doing Research in the Real World*. London: Sage.
- Green, P. & Skinner, D. (2005) Does time management training work? An evaluation, *International Journal of Training and Development*, 9(2), 124–139.
- Greene, J. C. (1988) Stakeholder participation and utilisation in program evaluation, *Evaluation Review*, 12, 91–116.
- Greene, J. C. & McClintock, C. (1991) The evolution of evaluation methodology, *Theory into Practice*, 30(1), 13–20.
- Greene, J. C., Benjamin, L. & Goodyear, L. (2001) The merits of mixing methods in evaluation, *Evaluation*, 7(1), 5–17.
- Gregory, A. (2000) Problematising participation, *Evaluation*, 6(2), 179–199.

- Grider, D. T., Capps, C. J., III & Toombs, L. A. (1990) Training evaluation: prescription for improvement, *Business*, 20–24.
- Griffin, R. (2011) Why trainers train but do not evaluate, *The Training Journal*, March, 57–59.
- Griffin, R. P. (2011) Workplace learning evaluation: a conceptual model and framework, *Industrial and Commercial Training*, 43(3), 172–178.
- Grondin, J. (1994). 'Introduction to Philosophical Hermeneutics'. Yale University Press.
- Gruenfeld, L. W. (1966) Personality Needs and Expected Benefits from a Management Development Programme, *Occupational Psychology*, 40, 75–81.
- Guba E. G. & Lincoln Y. S. (1994) Competing paradigms in qualitative research, *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publishers.
- Guba, E. G. & Lincoln, Y. S. (1989) Judging the quality of fourth generation evaluation, in Guba, E. G. & Lincoln, Y. S. (eds), *Fourth generation evaluation*, 5–21. Newbury Park, CA: Sage.
- Guba, E. G. & Lincoln, Y. S. (1989) *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- Guerci, M., Bartezzaghi, E. & Solari, L. (2010) Training evaluation in Italian corporate universities: a stakeholder-based analysis, *International Journal of Training and Development*, 14(4), 291–308.
- Guerci, M. & Vinante, M. (2011), Training evaluation: an analysis of the stakeholders' evaluation needs, *Journal of European Industrial Training*, 35(4), 385–410.
- Gummesson, E. (2000) *Qualitative Methods in Social Research*. Thousand Oaks, CA: Sage.
- Gutek, S. P. (1988) Training-program evaluation: an investigation of perceptions and practices in non-manufacturing business organisations (Doctoral dissertation). Kalamazoo, MI: Western Michigan University.
- Guyott, J. (1978) Management Training and Post-industrial Apologetics, *California Management Review*. 20(4), 84–93.
- Guzzo, R. A., Jette, R.D., Katzell, R.A., (1985). The Effects of Psychologically-Based Intervention Programs on Worker Productivity: A Meta-analysis, *Personnel Psychology* 38: 275–291.

- Haarich, S. N. (2005) Diferentes sistemas de evaluación de las políticas públicas en Europa: España, Alemania y los países del Este, *Revista Española de Control Externo*, 7(20), 61–87.
- Hall, B. (2003) Click here for ROI, *Training*, 40(3), 15.
- Hamblin, A. C. (1974) *Evaluation and Control of Training*. New York, NY: McGraw-Hill.
- Hammick M., Freeth D., Koppel I., Reeves S. & Barr, H. (2007) *A best evidence systematic review of interprofessional education: BEME Guide*, No. 9.
- Hanke, J. & Reitsch, A. (1992). Business Forecasting: Fourth Edition. New York, NY: Simon & Schuster.
- Hansen, L. P. (1982) Large sample properties of generalised method of moments estimators, *Econometrica*, 50, 1029–1054.
- Hansen, P. R. (2003) *A test for superior predictive ability* (Manuscript). Stanford, CA: Stanford University.
- Hansen, P. R., Lunde, A. & Nason, J. (2004) Model confidence sets for forecasting models (Manuscript). Stanford, CA: Stanford University.
- Hanson, R. (2011) Life after NPS, *Marketing Research*, Summer, https://www.marketstrategies.com/user_area/content_media/MR_Life_After_NPS.pdf [retrieved on 16 December 2014].
- Harari, Y. N. (2004) *Sapiens: A brief history of Humankind*. London: Harvill Secker Publishers.
- Hart, C. (1998) *Doing a Literature Review*. London: Sage.
- Hasan, U. & Nasreen, R. (2014) The Empirical Study of Relationship between Post-Purchase Dissonance and Consumer Behaviour, *Journal of Marketing Management*, 2(2), 65–77.
- Hashim, J. (2001), Training evaluation: Clients' Roles, *Journal of European Industrial Training*, 25(7), 374–379.
- Haslinda, A. & Mahyuddin, M. Y. (2009) The Effectiveness of Training in the Public Service, *American Journal of Scientific Research*, 6, 39–51.
- Haywood, K. M. (1992) Effective Training: Toward a Strategic Approach, *Cornell Hotel and Restaurant Administration Quarterly*, 33(4), 43–52.
- Hazer, J. T. & Highhouse, S. (1997) Factors Influencing Managers' Reactions to Utility Analysis: Effects of SOy Method, Information

- Frame, and Focal Intervention, *Journal of Applied Psychology*, 82, 104–112.
- Hazewinkel, M. (ed) (2001) Law of large numbers, *Encyclopedia of Mathematics*. Netherlands: Springer.
- Heath R. G., Brandt, D. & Nairn, A. (2006) Brand Relationships — Strengthened by Emotion, Weakened by Attention, *Journal of Advertising Research*. 46(4), 410–419.
- Heath, C. (1999) On the Social Psychology of Agency Relationships: Lay Theories of Motivation Overemphasise Extrinsic Incentives, *Organizational Behavior and Human Decision Processes*, 78(1), 25–62.
- Heath, R. G. & Feldwick, P. (2007) Fifty years using the wrong model of advertising, *International Journal of Market Research*, 50(1), 29–59 [http://www.bath.ac.uk/management/news_events/pdf/07-02-08.pdf].
- Heath, R. G. & Nairn, A. C. (2005) Measuring Emotive Advertising – Implications of Low Attention Processing on Recall, *Journal of Advertising Research*, 45(2), 269–281.
- Heath, R. G. (2001) The Hidden Power of Advertising, *Admap Monograph No. 7*. Henley-on-Thames: World Advertising Research Centre.
- Henning, E., Van Rensburg, W. & Smit, B. (2004) *Finding your way in qualitative research*. Pretoria: Van Schaik Publishers.
- Hermanson, R. (1964) Accounting for Human Assets, *Occasional Paper No. 14*. East Lansing, MI: Bureau of Business and Economic Research.
- Hertz, J., Krogh, A. & Palmer, R. (1990) *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.
- Heyes, J. (2000), Workplace Industrial Relations and Training, in Rainbird, H. (ed), *Training in the Workplace*. London: Macmillan.
- Hick, S. (2000) Why return on investment?, *Training and Development*. 54(7), 59.
- Hinkin, T. K. (1998) A brief tutorial on the development of measures for use in questionnaires, *Organisational Research Methods*, 1, 104–121.
- Hinton, G. E. & Anderson J. A. (eds). (1981) *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.

- Hoeksema, L. H., Van de Vliert, E. & Williams, A. R. T. (1997) The interplay between learning strategy and organisational structure in predicting career success, *International Journal of Human Resource Management*, 8, 307–327.
- Hogan, R. L. (2007) The historical development of program evaluation: exploring the past and present, *Online Journal of Workforce Education and Development*, 2(4), 1–14.
- Holly, L. & Rainbird, H. (2001) Workplace learning and the limits to evaluation, in Rainbird, H. (ed), *Training in the Workplace*, 264–282. London: Macmillan.
- Holte-McKenzie, M., Forde, S. & Theobald, S. (2006) Development of a participatory monitoring and evaluation strategy, *Evaluation and Program Planning*, 29, 365–376.
- Holton, E. (1996) The flawed four-level evaluation model, *Human Resource Development Quarterly*, 7(1), 5–21.
- Holton, E. (2005) Holton's evaluation model: new evidence and construct elaborations, *Advances in Developing Human Resources*, 7(1), 37–54.
- Holton, E. & Naquin, S. (2005) A critical analysis of HRD evaluation models from a decision-making perspective, *Human Resource Development Quarterly*, 16(2), 257–280.
- Holton, E. F., III, Reid, A., Bates, R. A. & Ruona, W. E. A. (2000) Development of a generalised learning transfer system inventory, *Human Resource Development Quarterly*, 11, 333–360.
- Horwitz, F. M. (1999) The emergence of strategic training and development: The current state of play, *Journal of European Industrial Training*, 23(4/5), 180–190.
- House, E. R. & Howe, K. R. (1999) *Values in Evaluation and Social Research*. Thousand Oaks, CA: Sage.
- Hu, L.-T. & Bentler, P. M. (1998) Fit indices in covariance structure modeling: sensitivity to underparameterised model misspecification, *Psychological Methods*, 3, 424–543.
- Hubbard, D. W. (2007) *How to measure anything: finding the value of intangibles in business*. London: Wiley & Sons.
- Hubbard, D. W. (2014) *How to measure anything: finding the value of intangibles in business. Volume 2* London: Wiley & Sons.
- Humphrey, C. (2014) *Predicting Future Lifespan: The Lindy Effect, Gott's Predictions and Caves' Corrections, and Confidence Intervals*. 2014

Joint Statistical Meeting, 2–7 August, Boston, MA, American Statistical Association, Alexandria, VA.

Hussey, D. (1988) *Management Training and Corporate Strategy: How to Improve Competitive Performance*. Oxford: Pergamon Press.

Hutchins, H. & Burke, L. (2007) Identifying trainers' knowledge of training research findings — closing the gap between research and practice, *International Journal of Training and Development*, 11(4), 236–264.

Ibrahim, A. (2008) *Evaluating training effectiveness in the Malaysian Public Service*, Durham Theses, Durham University, http://etheses.dur.ac.uk/2176/1/2176_184.PDF?UkUDh:CyT, [retrieved on 14 Dec 2014].

Illeris, K. (2003) Workplace learning and learning theory, *Journal of Workplace Learning*, 15(4), 167–178.

Jago, A. (1982) The Transfer of Learning to the Workplace, *Journal of European Industrial Training*, 6(7), 21–24.

James, C. & Roffe, I. (2000) The evaluation of goal and goal-free training innovation.

Jayawarna, D., Macpherson, A. & Wilson, A. (2007) Training Commitment and Performance in Manufacturing SMEs: Incidence, Intensity and Approaches, *Journal of Small Business and Enterprise Development*, 14(2), 321–338.

Jeffreys, H. (1961) *Theory of Probability*, 3rd edn. Oxford: Clarendon.

Jeffries, R., Miller, J. R., Wharton, C. & Uyeda, K. M. (1991) User interface evaluation in the real world: A comparison of four techniques, in Association for Computing Machinery (ed), *Proceedings of the ACM CHI 91 Conference on Human Factors in Computing Systems*, 119–124. New York, NY: ACM Press.

Jensen, M. C. (2003) Paying People to Lie: The Truth about the Budgeting Process, *European Financial Management*, 9(3), 379–406.

Jones, A. R., Griffin, M. A. & Rafferty, A. E. (2007) The executive coaching trend: towards more flexible executives, *Leadership and Organisation Development Journal*, 27(7), 584–596.

Kalemci, P. (2005) General Overview of Training Effectiveness and Measurement Models, *Journal of Commerce & Tourism Education Faculty*, 1, 144–156.

- Kanfer, R. & Ackerman, P. L. (1989) Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition, *Journal of Applied Psychology*, 74(4), 657–690.
- Kaplan, R. & Norton, D. (1996) *The balanced scorecard — translating strategy into action*. Boston, MA: Harvard Business School Press.
- Kaplan, R. S. & Norton, D. P. (1992) The balanced scorecard — measures that drive performance, *Harvard Business Review*, January/February, 71–79.
- Kaplan, R. S. & Norton, D. P. (1996a) Using the Balanced Scorecard as a strategic management system, *Harvard Business Review*, January/February, 75–85.
- Kaplan, R. S. & Norton, D. P. (1996b) Linking the Balanced Scorecard to Strategy, *California Management Review*, 39, 53–79.
- Kaplan, R. S. (2010) *Conceptual Foundations of the Balanced Scorecard*. Boston, MA: Harvard Business Press.
- Kaufman, R. & Keller, J. (1994) Levels of evaluation: beyond Kirkpatrick, *Human Resource Development Quarterly*, 5(4), 371–380.
- Kaufman, R., Keller, J. & Watkins, R. (1995) What works and what doesn't: evaluation beyond Kirkpatrick, *P&I*, 35(2), 8–12.
- Kayes, D. C. (2006) *Destructive goal pursuit: The Mount Everest disaster*. Basingstoke: Palgrave Macmillan.
- Kearns, P. (2000) *Maximising your ROI in training*. London: Financial Times/Prentice Hall.
- Kearns, P. (2005) *Evaluating the ROI from Learning*. CIPD.
- Kearsley, G. (1982) *Costs, Benefits, and Productivity in Training Systems*. Reading, MA: Addison-Wesley.
- Kerns, C. D. (2002) Managerial leadership at twelve o'clock: Link management practice to business results, *Graziadio Business Review*, 3, <http://gbr.pepperdine.edu/2010/08/managerial-leadership-at-twelve-oclock/> [retrieved on 31 May 2011].
- Kerns, C. D. (2005) *Value – Centered Ethics: A Proactive System to Shape Ethical Behavior*. Amherst, MA: HRD Press.
- Kerns, C. D. & Ko, K., (2010) Exploring happiness and performance at work, *Journal of Organisational Leadership & Business*, Summer, 1–24.

- Kerns C. D. & Po, K. (2012) An alternative to assessing ROI in 'softer' organisational initiatives, *Journal of Organisational Leadership & Business*, Fall.
- Kerr, S. (1975) On the folly of rewarding A, while hoping for B, *The Academy of Management Journal*, 18(4), 769–783.
- Kerr, S. (1995) On the folly of rewarding A, while hoping for B (Update). *The Academy of Management Executive*, 9(1), 7–14.
- Khalid, M., Rehman, C. & Ashraf, M. (2012) Exploring the link between Kirkpatrick (KP) and context, input, process and product (CIPP) training evaluation models, and its effect on training evaluation in public organisations of Pakistan, *African Journal of Business Management*, 6(1), 274–279.
- King, J. A. (2007) Making sense of participatory evaluation, *New Directions for Evaluation*, 114, 83–105.
- King, L. A. & Burton, C. M. (2003) The hazards of goal pursuit, in Chang, E. C. & Sanna, L. J., *Virtue, vice, and personality: The complexity of behavior*. Washington, DC: American Psychological Association.
- Kirkpatrick, D.L. (1954). *Evaluating Human Relations Programs for Industrial Foremen and Supervisors*. PHD Thesis. Kirkpatrick Publishing.
- Kirkpatrick, D. L. (1976) Evaluation of Training, *Training and development handbook: A guide to human resource development*. New York, NY: McGraw-Hill Company.
- Kirkpatrick, D. L. (1977) Evaluating training programs: Evidence vs proof, *Training and Development Journal*, 9–12.
- Kirkpatrick, D. L. (1998) *Evaluating Training Programs: The Four Levels*, 2nd edn. San Francisco, CA: Berrett-Koehler Publishers.
- Kirkpatrick, D. (1979) Techniques for evaluating training programs, *Training and Development Journal*, 78–92.
- Kirton, M. J. (1978) Field Dependence and Adaptation Innovation Theories, *Perceptual and Motor Skills*, 47, 1239–1245.
- Kirton, M. J. (2003) *Adaptation and innovation in the context of diversity and change*. London: Routledge.
- Kirwan, C. & Birchall, D. (2006) Transfer of learning from management development programmes: testing the Holton model, *International Journal of Training and Development*, 10(4), 252–268.

- Knight, D., Durham, C. C. & Locke, E. A. (2001) The Relationship of Team Goals, Incentives, and Efficacy to Strategic Risk, Tactical Implementation, and Performance, *The Academy of Management Journal*, 44(2), 326–338.
- Kohonen, T. (1977) *Associative memory: A system theoretical approach*. New York, NY: Springer-Verlag.
- Kontogiorghe, C. (2001) Factors affecting training effectiveness in the context of the introduction of a new technology – a US case study, *International Journal of Training and Development*, 5(4), 248–260.
- Kontogiorghe, C. (2004) Reconceptualising the learning transfer conceptual framework: empirical validation of a new systemic model, *International Journal of Training and Development*, 8(3), 210–221.
- Kraiger, K., Ford, J. K. & Salas, E. (1993) Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation, *Journal of Applied Psychology*, 78(2), 311–328.
- Kruger, J. & Dunning, D. (1999) Unskilled and unaware of it: How differences in recognising one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
- Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*, 2nd edn. Chicago, IL: University of Chicago Press.
- Kumpikaite, V. (2007) Human resource training evaluation, *Engineering Economics*, 5(55), 30–36.
- Landers, R. N. and Callan R. C. (2012) Training Evaluation in Virtual Worlds: Development of a Model Managerial and Commercial Applications, *Journal of Virtual Worlds Research*, 5(3), 1.
- Larrick, R. P., Heath, C. & Wu, G. (2009) Goal-Induced Risk Taking in Negotiation and Decision Making, *Social Cognition*, 27, 342–364.
- Latham, G. P. (1988) Human Resource Training and Development, *Annual Review of Psychology*, 75, 545–582.
- Latham, G. P. (1989) Behavioural approaches to training and learning processes, in Goldstein, I. L. (ed), *Training and Development in Organisations*, 256–295, San Francisco, CA: Jossey Bass.
- Latham, G. P. & Locke, E. A. (2006) Enhancing the Benefits and Overcoming the Pitfalls of Goal Setting, *Organisational Dynamics*, 35(4), 332–340.

- Latham, G. P. & Whyte, G. (1994) The futility of Utility analysis, *Personnel Psychology*, 47.
- Lave, J. & Wenger, E. (1991) *Situated Learning – Legitimate Peripheral Participation*. Cambridge, MA: Cambridge University Press.
- Lavidge, R. J. & Steiner, G. A. (1961) A Model for Predictive Measurements of Advertising Effectiveness, *Journal of Marketing*, 25(4), 59–62.
- Lawrence, M., Goodwin, P., O'Connor, M. & Onkal, D. (2006) Judgemental Forecasting: A Review of Progress Over the Last 25 Years, *International Journal of Forecasting*, 22, 493–518.
- Lazarsfeld, P. F. & Rosenberg, M. (1955) *The Language of Social Research*. New York, NY: The Free Press.
- Learning and Skills Council. (2009) *National Employers Skills Survey 2008*. Coventry: Learning and Skills Council.
- Lee, S. H. & Pershing, J. A. (2002) Dimensions and design criteria for developing training reaction evaluations, *Human Resource Development International*, 5(2), 175–197.
- Lee, Y, S. (2007) An Approach to Identify Effective Learning Outcomes for a Training Program (Master of Science Thesis). Blacksburg, VA: Virginia Polytechnic Institute and State University.
- Leedham, M. (2005) The Coaching scorecard: A holistic approach to evaluating the benefits of business coaching, *International Journal of Evidence-Based Coaching and Mentoring*, 3(2), 30–41.
- Leimbach, M. (2010) Learning transfer model: a research driven approach to enhancing learning effectiveness, *Industrial and Commercial Training*, 42(2), 81–86.
- Lewis, P. & Thornhill, A. (1994) The evaluation of training: An Organisational Approach, *Journal of European Industrial Training*, 18(8), 25–33.
- Lewis, T. (1996) A model for thinking about the evaluation of training, *Performance Improvement Quarterly*, 9(1), 3–22.
- Lieberson, S. (2000) Small N's and big conclusions: an examination of the reasoning in comparative studies based on a small number of cases, in Gomm, R., Hammersley, M. & Forster, P. (eds), *Case Study Method: Key Issues, Key Texts*. London: Sage.
- Likert, R. (1961) *New Patterns of Management*. New York, NY: McGraw-Hill.

- Lincoln, Y. (2001) Engaging sympathies: relationships between action research and social constructivism, in Reason, P. & Bradbury, H. (eds) *The SAGE handbook of action research: participative inquiry and practice*, 124–132. Thousand Oaks, CA: Sage.
- Lingham, T., Richley, B. & Rezania, D. (2006) An evaluation system for training programs: A case study using a four-phase approach, *Career Development International*, 11(4), 334–351.
- Little, B. (2004) Best practices to ensure the maximum ROI in L&D, *Industrial and Commercial Training*, 46(7), 400–405.
- Locke, E. A. & Latham, G. P. (2002) Building a practically useful theory of goal setting and task motivation: a 35-year odyssey, *American Psychologist*, 57(9), 705–717.
- Locke, E. A. & Latham, G. P. (2006) New Directions in Goal-Setting Theory, *Current Directions in Psychological Science*, 15(5), 265–268.
- Locke, E. A., Latham, G. P., Smith, K. J. & Wood, R. E. (1990) *A Theory of Goal Setting & Task Performance*. Boston, MA: Prentice Hall College Div.
- Lucas, R. E. (1976). Econometric policy evaluation: A critique. In: Lucas R E Jr (ed.) *Studies in Business Cycle Theory*. MIT Press, Cambridge, MA (originally published in 1976).
- Luoma, M. (2005) Managers' perception of the strategic role of management development, *Journal of Management Development*, 24(7), 645–655.
- Luthans, F., Avey, J. B., Avolio, B. J., Norman, S. M. & Combs, G. J. (2006) Psychological capital development: toward a micro-intervention, *Journal of Organizational Behavior*, 27, 387–393.
- Luthans, F., Youssef, C. M. & Avolio, B. J. (2007) *Psychological Capital: Developing the Human Competitive Edge*. New York, NY: Oxford University Press.
- Mabey, C., Salaman, G. & Storey, J. (1998) *Strategic human resource management: a reader*. London: Sage.
- MacInnis, D. J. & Jaworski, B. J. (1989) Information processing from advertisements: towards an integrative framework, *Journal of Marketing*, 3(4) 1–23.
- Madaus, G.F., Scriven, M. S. & Stufflebeam, D. L. (1986) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston, MA: Kluwer-Nijhoff.

- Maisel, L. S. & Cokins, G. (2014) *Predictive Business Analytics*. San Fransisco, CA: Wiley.
- Makridakis, S. (1986) The Art and Science of Forecasting, *International Journal of Forecasting*, 2, 45.
- Makridakis, S. G., Wheelwright, S. C. & Hyndman, R. J. (1998) *Forecasting: Methods and Applications*, 3rd edn. New York: John Wiley & Sons.
- Mandelbrot, B. B. (1984) *The fractal geometry of Nature*. New York, NY: Freeman.
- Mann, L., Samson, D. & Dow, D. (1998) A Field Experiment on the Effects of Benchmarking and Goal Setting on Company Sales Performance, *Journal of Management*, 24(1), 73–96.
- Mann, S. (1996) What should Training Evaluations evaluate?, *Journal of European Industrial Training*, 20(9), 14–20.
- Marjanovic, B. (2016) An (Old) Way to Pick (New) Stocks, *Seeking Alpha*, <http://seekingalpha.com/article/3834026-old-way-pick-new-stocks> [Retrieved 27 January 2016].
- Mark, M. M. & Shotland, R. L. (1985) Stakeholder-based evaluation and value judgments, *Evaluation Review*, 9, 605–626.
- Mark, M. M., Henry, G. T. & Julnes, G. (2000) *Review of Evaluation: An Integrated Framework for Understanding, Guiding and Improving Policies and Programs*. San Francisco, CA: Jossey Bass.
- Massey, C. (2004) Is the training train out of control? A case of evaluation failure from New Zealand, *Journal of Small Business and Enterprise Development*, 11(4), 458–466.
- Mat, K., Omar, M. Z., Osman, S. A., Kofli, N. T., Rahman, M. N., Abd., Jamil, M. & Jamaluddin, N. (2011) The Effectiveness of Industrial Training on UKM Engineering Students, *Procedia Social and Behavioral Sciences*, 18, 656–665.
- Mathie, A. & Greene, J. C. (1997) Stakeholder participation in evaluation: how important is diversity?, *Evaluation and Program Planning*, 20, 279–285.
- Maxwell, J. A. (1996) *Qualitative Research Design: An Interactive Approach*. Thousand Oaks, CA: Sage.
- Mayo, E. (1946) *Human Problems of an Industrial Civilisation*, London: Macmillan.

- Mayne, J. and Rist, R.C. (2006). Studies are not enough: the necessary transformation of evaluation. *Canadian Journal of Program Evaluation* 21: 93-120.
- McClelland, J. L. (1981) Retrieving general and specific information from stored knowledge of specifics, in Cognitive Science Society (eds), *Proceedings of the Third Annual Conference of the Cognitive Science Society*. Berkeley, CA: Cognitive Science Society.
- McClelland, J. L. & Rumelhart, D. E. (1985) Distributed memory and the representation of general and specific information, *Journal of Experimental Psychology: General* 114, 159–188.
- McDaniel, M., Schmidt, F. L. & Hunter, J. E. (1987) Job experience as a determinant of job performance: *95th Annual Convention of the American Psychological Association*, August, New York, NY, American Psychological Association, Washington, DC.
- McGovern, M., Lindemann, M., Vergara, M., Murphy, S., Barker, L. & Warrenfeltz, R. (2001) Maximising the impact of executive coaching: behavioral change, organisational outcomes, and return on investment, *Manchester Review*, 6(1), 1–7, www.coachfederation.org/includes/docs/049ManchesterReviewMaximisingImpactofExecCoaching2.pdf [retrieved 29 January 2012].
- McGuire, D., Garavan, T. N., O'Donnell, D., Saha, S. K. & Cseh, M. (2008) Managers' personal values as predictors of importance attached to training and development: a cross country exploratory study, *Human Resource Development International*, 11(4), 335–350.
- McLean, G. N. (2005) Examining approaches to HR evaluation: the strengths and weaknesses of popular measurement methods, *Strategic Human Resources*, 4(2), 24–27.
- McLinden, D. J. (1995) Proof, evidence, and complexity: understanding the impact of training and development in business, *Performance Improvement Quarterly*, 8, 3–18.
- McLinden, D. J. & Trochim, W. M. K. (1998) Getting to parallel: assessing the return on expectations of training, *Performance Improvement*, 37, 21–26.
- McNamara, D. S., Kintsch, E., Songer, N. B. & Kintsch, W. (1996) Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text, *Cognition & Instruction*, 14, 1–43.

- Meehl, P. E. (1978) Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald and the Slow Progress of Soft Psychology, *Journal of Consulting and Clinical Psychology* 46, 817.
- Meese, R. A., Rogoff, K. (1983) Empirical exchange rate models of the seventies: Do they fit out of sample?, *Journal of International Economics*, 14, 3–24.
- Meese, R. A., Rogoff, K. (1988) Was it real? The exchange rate — interest differential over the modern floating rate period, *Journal of Finance*, 43, 933–948.
- Meny, Y. & Thoenig, J. C. (1989) *Politiques Publiques*. Paris: PUF.
- Merton, R. K. (1967) *On Theoretical Sociology*. New York, NY: New York Free Press.
- Michalski, G. V. & Cousins, J. B. (2000) Differences in stakeholder perceptions about training evaluation: a concept mapping/pattern matching investigation, *Evaluation and Program Planning*, 23, 211–230.
- Michalski, G. V. & Cousins, J. B. (2001) Multiple perspectives on training evaluation: probing stakeholder perceptions in a global network development firm, *American Journal of Evaluation*, 22(1), 37–53.
- Middleton, M. & Midgley, C. (1997) Avoiding the demonstration of lack of ability: An unexplored aspect of goal theory, *Journal of Educational Psychology*, 89, 710–718.
- Miles, M. B. & Huberman, A. M. (1994) *Qualitative Data Analysis*, 2nd edn. Thousand Oaks, CA: Sage.
- Milgram, P. & Kishino, F. (1994) A taxonomy of mixed reality visual displays, *IEICE Transactions on Information Systems*, E77-D(12), 1–15.
- Miller, T. I. (1994) Designing and conducting surveys, in Wholey, J. S., Hatry, H. P. & Newcomer, K. E. (eds), *Handbook of Practical Program Evaluation*, 271–292. San Francisco, CA: Jossey Bass.
- Millett, S. & Honton, E. (1991) *A Manager's Guide to Technology Forecasting and Strategy Analysis Methods*. New York, NY: Battelle Press.
- Mitchell, T. R. & Silver, W. S. (1990) Individual and group goals when workers are interdependent: Effects on task strategies and performance, *Journal of Applied Psychology*, 75(2), 185–193.

- Misrach, B. (1995) *Forecast comparison in L2* (Manuscript). New Brunswick, NJ: Rutgers University.
- Mmobuosi, L. (1987) Resolving Re-entrants' Problems in the Transfer of Management Learning, *Journal of European Industrial Training*, 11(1), 13–16.
- Modis, T. (1992) *Predictions: Society's Telltale Signature Reveals the Past and Forecasts the Future*. New York, NY: Simon & Schuster.
- Molich, R. & Nielsen, J. (1990) Improving a human-computer dialogue, *Communications of the ACM*, 33(3), 338–348.
- Monroe, K. B. & Krishnan, R. (1985) The Effect of Price on Subjective Product Evaluations, in Jacoby, J. & Olsen, J. (eds), *Perceived Quality*. Lexington, MA: Lexington Books.
- Mooi, L. M. (2010) Teacher Education: Innovative, Effectiveness and Global, *Procedia Social and Behavioral Sciences*, 9, 1815–1820.
- Morgan, W. A. (1939) A test for significance of the difference between two variances in a sample from a normal bivariate population, *Biometrika* 31, 13–19.
- Mossholder, K. W. (1980) Effects of externally mediated goal setting on intrinsic motivation: A laboratory experiment, *Journal of Applied Psychology*, 65(2), 202–210.
- Mulvaney, R. R. H., Zwahr, M. & Baranowski, L. (2006) The trend towards accountability: What does it mean for HR Managers?, *Human Resource Management Review*, 16, 431–422.
- Mumford, A. (1993) *How Managers Can Develop Managers*. Aldershot: Gower.
- Mussweiler, T. & Strack, F. (2000) The 'relative self': Informational and judgmental consequences of comparative self-evaluation, *Journal of Personality and Social Psychology*, 79(1), 23–38.
- Muth, J. (1961) Rational Expectations and the Theory of Price Movements, *Econometrica*, 29, 315–335.
- Myers, J. L. (1966) *Fundamentals of experimental design*. Boston, MA: Allyn and Bacon, Inc.
- Nan, S. A. (2003) Formative Evaluation, in Burgess, G. & Burgess, H. (eds), *Beyond Intractability*. Boulder, CO: Conflict Information Consortium, University of Colorado.

- National Skills Taskforce. (2000) *Third Report of the National Skills Taskforce: Tackling the Adult Skills Gap: Upskilling*. London, Department for Education and Employment.
- Neale, M. A. & Bazerman, M. H. (1985) The Effect of Externally Set Goals on Reaching Integrative Agreements in Competitive Markets, *Journal of Occupational Behaviour*, 6(1), 19–32.
- Neisser, U. (1979) The Concept of Intelligence, *Intelligence*, 3(3), 217–227.
- Newstrom, J. W. (1987) Anomalies in Evaluation, *Training and Development Journal*, July, 55–59.
- Nickols, F. W. (2005) Why a stakeholder approach to evaluation training?, *Advances in Developing Human Resources*, 7(1), 121–134.
- Nielsen, J. (1994) Heuristic evaluation, in Nielsen, J. & Mack, R. L. (eds), *Usability Inspection Methods*. New York, NY: John Wiley & Sons.
- Nielsen, J. (1990) Paper versus computer implementations as mockup scenarios for heuristic evaluation: *Third International Conference on Human-Computer Interaction*, 27–31 August, Cambridge, International Federation for Information Processing, Laxenburg.
- Nielsen, J. (1992) Finding usability problems through heuristic evaluation: *ACM CHI '92 Conference on Human Factors in Computing Systems*, 3–7 May, Monterey, CA, Association for Computing Machinery, New York, NY.
- Nielsen, J. & Landauer, T. K. (1993) A mathematical model of the finding of usability problems: *INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, 24–29 April, Amsterdam, Association for Computing Machinery, New York, NY.
- Nielsen, J. & Molich, R. (1990) Heuristic evaluation of user interfaces: *ACM CHI '90 Conference on Human Factors in Computing Systems*, 1–5 April, Seattle, WA, Association for Computing Machinery, New York, NY.
- Noe, R. A. (1986) Trainees' Attributes and Attitudes: Neglected Influence on Training Effectiveness, *Academy of Management Review*, 11, 736–749.
- Oppenheim, A. N. (1992) *Questionnaire Design, Interview and Attitude Measurement*. London: Pinter.

- Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D. & Bazerman, M. H. (2009) *Goals Gone Wild: The Systematic Side Effects of Over-Prescribing Goal Setting*. Boston, MA: HBR Press.
- Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Ostrom, E. (2000) Collective action and the evolution of social norms, *Journal of Economic Perspectives*, 14(3), 137–158.
- Ostrom, E., Gardner, R. & Walker, J. (1994) *Rules, Games and Common-pool Resources*. Ann Arbor, MI: University of Michigan Press.
- Paisey, C., Paisey, N. J. & Tarbet, H. (2007) Continuing professional development activities of UK accountants in public practice, *Accounting Education*, 16(4), 379–403.
- Paquet, B., Kasl, E., Weinstein, L. and Waite, W. (1987) The Bottom Line of Management Training, *Training and Development Journal*, 41(5), 27–34.
- Paradise, A. & Patel, L. (2009) *State of the Industry Report*. San Francisco, CA: American Society of Training and Development.
- Park, S. H. & Luo, Y. (2001) Guanxi and organisational dynamics: organisational networking in Chinese firms, *Strategic Management Journal*, 22, 455–477.
- Passmore, J. & Velez, M. (2012), SOAP-M: a training evaluation model for HR, *Industrial and Commercial Training*, 44(6), 315–325.
- Pastor, J. C. & Mayo, M. (2008) Transformational leadership among Spanish upper echelons: the role of managerial values and goal orientation, *Leadership & Organisation Development Journal*, 29(4), 340–358.
- Patterson, F. & Hobley, S. (2003) Measuring application, not just acquisition, *Journal of Knowledge Management*, 6(3).
- Patton R. J. & Lui G. P. (2014) *Eigenstructure assignment for control control systems, robotics and automation*, Vol. VIII. Paris: UNESCO.
- Patton, M. (2002) *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage Publications. (Quoted in: Creswell, 2007: 22.)
- Patton, M. Q. (1994) Developmental Evaluation, *Evaluation Practice*, 15(3), 311–319.

- Patton, M. Q. (1997) *Utilisation-Focused Evaluation: The New Century Text*, 3rd edn. Thousand Oaks, CA: Sage.
- Patton, M. Q. (1997) *Utilisation-focused evaluation: The new century text*. Thousand Oaks, CA: Sage.
- Patton, M. Q. (1998) *Utilisation-focused Evaluation* (3rd edn). Beverly Hills, CA: Sage.
- Pawson, R. & Tilley, N. (1997) *Realistic evaluation*. London, UK: Sage.
- Pellicer, E., Yepes, V. & Rojas, R. (2010) Innovation and competitiveness in construction companies — A case study, *Journal of Management Research*, 10(2), 103–115.
- Perks, J. & Bar-On, R. (2009) Coaching for emotionally intelligent leadership, in Passmore, J. (ed), *Leadership in Coaching: Working with Leaders to Develop Elite Performance*, 55–74. London: Kogan Page.
- Peters, O. (1994) Distance education in a post-industrial society, in Keegan, D. (ed), *Otto Peters on Distance Education*, 220–240. London: Routledge.
- Philips, J. J. (1991) *Handbook of Training Evaluation and Measurement Methods*. Houston, TX: Gulf Publishing Company.
- Phillips, J. J. (1996) *Accountability in Human Resource Management: Techniques for Evaluating the Human Resource Function and Measuring its Bottom-Line Contribution*. Houston, TX: Gulf Publishing Company.
- Philips, J. J. (1996) How much is the training worth?, *Training and Development*, 50(4), 20–24.
- Phillips, J. J. (1996) ROI: the search for best practices, *Training and Development*, 50(2), 42–47.
- Phillips, J. J. (1997) *Return on Investment in Training and Performance Improvement Programs*. Houston, TX: Gulf Publishing Company.
- Phillips, J. J. (2000) *The Consultant's Scorecard: Tracking Results and Bottom-Line Impact on Consulting Projects*. New York, NY: McGraw-Hill.
- Phillips, J. J. (2002) *Return on Investment in Training and Performance Improvement Programs*, 2nd edn. Woburn: Butterworth-Heinemann.
- Phillips, J. J. (2007) Measuring ROI: Fact, fad, or fantasy?, *T+D*, 61(1), 42–46.

- Phillips, S. & Edwards, R. (2000). Development, impact assessment and the praise culture, *Critique of Anthropology*, 20(1), 47–66.
- Plant, R. A., Ryan, R. J. (1992) Training Evaluation: A Procedure for validating an organisation's investment in training, *Journal of European Industrial Training*, 16(10), 22–38.
- Pohl, F. (1993) The Uses of the Future, *The Futurist*, March–April, 9.
- Polanyi, K. (1944) *The Great Transformation*. New York, NY: Holt, Rinehart & Winston Inc.
- Polit, D. F. & Beck, C. T. (2008) *Nursing Research; Generating and Assessing Evidence for Nursing Practice*, 8th edn. Philadelphia, PA: Lippincott Williams & Wilkins.
- Preskill, H. & Torres, R. T. (1999) Building Capacity for Organisational Learning Through Evaluative Inquiry, *Evaluation*, 5(1), 41–60.
- Pulley M.L. (1994) 'Navigating the Evaluation Rapids' *Training and Development*. Vol 48(9), September, pp. 19-24
- Punia, B. K. & Kant, S. (2013) A Review of factors affecting training effectiveness vis-à-vis managerial implications and future research, *Directions — International Journal of Advanced Research in Management and Social Sciences*, 2(1), 151–164.
- Purcell, A. (2000) 20/20 Return on investment, *Training and Development*, 54(7), 29–33.
- Quesada, C., Pilar P. H. & Berta, E. (2011) Evaluating the Efficiency of Leadership Training Programmes in Spain, *Procedia — Social and Behavioral Sciences*, 30, 2194–2198.
- Raab, R. T., Swanson, B. E. & Wentling, T. L. (1991) *Improving Training Quality*. Rome: Food and Agriculture Organisation of the United Nations.
- Raab, R. T., Swanson, B. E., Wentling, T. L. & Dark, C. D. (eds) (1991) *A trainer's guide to evaluation*. Rome: FAO.
- Rae, L. (1983) End of Course Validation, *The Training Officer*, October, 306–310.
- Rama, M. J. & Vaishnavi. R. (2012) Measuring Training Effectiveness: A Study in a Leading Retailing Industry in A Metropolitan City, *International Journal of Multidisciplinary Management Studies*, 2(4), 16–25.

- Raskin, J. D. (2002) Constructivism in psychology: Personal construct psychology, radical constructivism, and social constructivism, In Raskin, J. D. & Bridges, S. K. (eds), *Studies in meaning: Exploring constructivist psychology*, 1–25. New York, NY: Pace University Press.
- Rawsthorne, L. J. & Elliot, A. J. (1999) Achievement Goals and Intrinsic Motivation: A Meta-Analytic Review, *Personality and Social Psychology Review*, 3(4), 326–344.
- Rayner, K. & Serano, S. C. (1994) Eye movements in reading: Psycholinguistic Studies, in Gernsbacher, M. A. (ed), *Handbook of Psycholinguistics*. San Diego, CA: Academic Press.
- Reason, P. & Bradbury, H. (2001) *Handbook of Action Research: Participative Inquiry and Practice*. London: Sage.
- Reichheld, F. (2003) One Number You Need to Grow, *Harvard Business Review*, 81(12), 46–54.
- Reid, M. A. & Barrington, H. (1999) *Training Interventions: Promoting Organisational Learning*, 6th edn. London: Chartered Institute of Personnel Development.
- Reilly, R. R. & Smither, J. W. (1983) An examination of two alternative techniques to estimate the standard deviation of job performance in dollars, *Journal of Applied Psychology*, 70, 651–661.
- Reineke, R. (1991) Stakeholder involvement in evaluation: suggestion for practice, *American Journal of Evaluation*, 12(39), 39–44.
- Rickards, T., Chen, M. & Moger, S. (2001) Development of a self-reported instrument for exploring team factor, leadership and performance relationships, *British Journal of Management*, 12, 243–250.
- Roark, S., Kim, M. & Mupinga, M. (2006) An exploratory study of the extent to which medium-sized organisations evaluate training programs, *Journal of Business and Training Education*, 15, 15–20.
- Robertson, M. (2004) Building program theory for evaluation: the process through a political lens (Unpublished Doctoral dissertation). Athens, GA: University of Georgia.
- Robinson, G. F., Erlen, J. A., Rubio, D. M., Kapoor, W. N. & Poloyac, S. M. (2013) Development, implementation, and evaluation of an interprofessional course in translational research, *Journal of Clinical and Translational Science*, 6(1), 50–56.
- Robotham, D. (1995) Self-Directed Learning: The Ultimate Learning Style?, *Journal of European Industrial Training*, 19(7), 3–7.

- Rock, D. (2008). SCARF: A brain-based model for collaborating with and in influencing others. *NeuroLeadership Journal*, 1, 44–52.
- Rodríguez-Bilella, E. P. & Monterde-Díaz, R. (2012) Valuation, Valuation, Negotiation: Some Reflections towards a Culture of Evaluation, *The Canadian Journal of Program Evaluation*, 25(3), 1–10.
- Roedel, T. D., Schraw, G. & Plake, B. S. (1994) Validation of a measure of learning and performance goal orientation, *Educational and Psychological Measurement*, 54(4), 1013–1021.
- Ross, J. A. (2008) Cost-utility analysis in educational needs assessment, *Evaluation and Program Planning*, 31, 356–367.
- Rossi, P. H., Freeman, H. E. & Lipsey, M. W. (1999) *Evaluation: A Systematic Approach*, 6th edn. Thousand Oaks, CA: Sage Publications.
- Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004) *Evaluation: A Systematic Approach*, 7th edn. London: Sage Publications.
- Rouiller, J. Z. & Goldstein, I. L. (2006) The relationship between organisational transfer climate and positive transfer of training, *Human Resource Development Quarterly*, 4(4), 377–390.
- Rowden, R. W. (2005) Exploring methods to evaluate the return-on-investment from training. *Business Forum*, 27(1), 31–36.
- Rowley, T. J., Behrens, B. & Krackhardt, D. (2000) Redundant governance structures: an analysis of structural and relational embeddedness in the steel and semiconductor industries, *Strategic Management Journal*, 21, 369–386.
- Rumelhart, D. E., Hinton, G. E. & McClelland, J. L. (1986) A general framework for parallel distributed processing, in Rumelhart D. E., McClelland, J. L. & the PDP Research Group (eds), *Parallel distributed processing: Explorations in the microstructures of cognition*, Vol. 1. Cambridge, MA: MIT Press.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986) Learning internal representations by error propagation, in Rumelhart, D. E., McClelland, J. L. & the PDP Research Group (eds), *Parallel distributed processing: Explorations in the microstructures of cognition*, Vol. 1. Cambridge, MA: MIT Press.
- Rumelhart, D. E., McClelland, J. L. & the PDP Research Group (1986) *Parallel distributed processing: Explorations in the microstructure of cognition*, Vols. 1 & 2. Cambridge, MA: MIT Press.

- Russ-Eft, D. & Preskill, H. (2008) Improving the quality of evaluation participation: A meta-evaluation, *Human Resource Development International*, 11(1), 35–50.
- Russ-Eft, D., Bober, M., Teja, I., Foxon, M. & Koszalka, T. (2008) *Evaluator Competencies: Standards for the Practice of Evaluation in Organizations*. San Francisco, CA: John Wiley & Sons, 3–12.
- Russell, B. (1912) *The Problems of Philosophy*. New York, NY: Oxford University Press.
- Sahinidis, A. G. & Bouris, J. (2008) Employee Perceived Training Effectiveness Relationship to Employee Attitudes, *Journal of European Industrial Training*, 32(1), 63–76.
- Saks, A. & Haccoun, R. (2007) *Managing Performance through Training and Development*. Toronto: Nelson and Thompson Ltd.
- Salas, E. & Cannon-Bowers, J. A. (2001) The science of training: a decade of progress, *Annual Review of Psychology*, 52, 471–497.
- Sanderson, G. (1995) Objectives and evaluation, in Truelove, S. (ed), *Handbook of Training and Development*, 2nd edn. Oxford: Blackwells.
- Santos, A. & Stuart, M. (2003) Employee Perceptions and their influence on Training Effectiveness, *Human Resource Management Journal*, 13(1), 27–45.
- Schmidt, F. L., Hunter, J. E. & Pearlman, K. (1982) Assessing the economic impact of personnel programs on workforce productivity, *Personnel Psychology*, 35, 333–347.
- Schmidt, S. W. (2009) Employee demographics and job training satisfaction: the relationship between dimensions of diversity and satisfaction with job training, *Human Resource Development International*, 12(3), 297–312.
- Schon, D. A. (1983) *The Reflective Practitioner*. New York, NY: Basic Books.
- Schraw, G. & Dennison, R. S. (1994) Assessing metacognitive awareness, *Contemporary Educational Psychology*, 19, 460–475.
- Schraw, G. & Moshman, D. (1995) Metacognitive Theories, *Educational Psychology Review*, 7(4), 351–371.
- Schwandt, T. A. (1994) Constructivist, interpretive approaches to human inquiry, *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publishers.

- Schwandt, T. A. (1992) Better Living Through Evaluation? Images of progress shaping evaluation practice, *Evaluation Practice*, 12(3), 135–144.
- Schweitzer, M. E., Ordóñez, L. & Douma, B. (2004) Goal Setting as a Motivator of Unethical Behavior, *Academy of Management Journal*, 47(3), 422–432.
- Scourtoudis, L. & Dyke, L. (2007) Assessing the behavioural change and organisational outcomes resulting from management training, *International Journal of Learning*, 13(10), 75–84.
- Schmalenbach, M. (2005). *Training Evaluation* ©TrainingZONE Pub January 2005. Downloaded July 3rd 2013
- Scriven, M. (1990) The evaluation of hardware and software, *Studies in Educational Evaluation*, 16, 3–40.
- Scriven, M. (1991) Beyond Formative and Summative Evaluation, in McLaughlin, M. W. & Phillips, E. D. C. (eds), *Evaluation and Education: A Quarter Century*. Chicago: University of Chicago Press.
- Scriven, M. (1993) *Evaluation Thesaurus*. London: Sage.
- Scriven, M. (1994) Product evaluation: The state of the art, *Evaluation Practice*, 15(1), 45–62.
- Scriven, M. (1996) Goal-free evaluation, *Evaluation News and Comment Magazine*, 5(2), 5–9.
- Seigal, E. (2013). *Predictive Analytics*. Pub: John Wiley & Sons Inc. Hoboken New Jersey.
- Senensky, B. (2008) Predictive Modeling, *CompAct (Society of Actuaries, Technology Section newsletter)*, July, 14.
- Seppanen-Jarvela, R. (2005) Internal evaluation of a management-development initiative: a public-sector case, *The Journal of Management Development*, 24(1), 45–56.
- Shadish, W. R. (1992) Theory-driven meta-evaluation, in Chen, H.-T. & Rossi, P. H. (eds), *Using Theory to Improve Program and Policy Evaluation*, Westport, CT: Greenwood Press.
- Shadish, W. R., Cook, T. D. & Leviton, L. C. (1991) *Foundation of Program Evaluation*. Beverly Hills, CA: Sage.
- Shalley, C. E. & Oldham, G. R. (1985) Effects of Goal Difficulty and Expected External Evaluation on Intrinsic Motivation: A Laboratory Study, *The Academy of Management Journal*, 28(3), 628–640.

- Shamsuddin, A. (1995) Contextual factors associated with evaluation practices of selected adult and continuing education providers in Malaysia (Unpublished Doctoral dissertation). Athens, GA: University of Georgia.
- Shani, A. B., Mohrman, S. A., Pasmore, W. A., Stymne, B. & Adler, N. (2007) *Handbook of Collaborative Research*. London: Sage.
- Shaw, R. & Merrick, D. (2005) *Marketing Payback: Is your Marketing Profitable?*. Harlow: Financial Times Prentice Hall
- Shaw, R. B. & Perkins, D. N. T. (1991) Teaching Organisations to Learn, *Organisation Development Journal* 9(4), 1–12.
- Shelton, S. & Alliger, G. (1993) Who's afraid of Level 4 Evaluation?, *Training & Development Journal*, 47(6), 43–46.
- Shepherd, C. (1999) Assessing the ROI of Training, www.fastrak-consulting.com.uk [retrieved on March 23, 2012].
'https://www.academia.edu/17502221/Assessing_the_ROI_of_training'.
- Sherman, S. (1995) Stretch goals: The dark side of asking for miracles, *Fortune*, 132(10), 231.
- Shiv, B. & Fedorikhin, A. (1999) Heart & Mind in Conflict: The interplay of Affect and Cognition in Consumer Decision Making, *Journal of Consumer Research*, 26, 278–292.
- Short, T. (2009) Exploring the vacuum in training evaluation: Is this a case of mission impossible?, *Development and Learning in Organisations*, 23(5), 15–18.
- Siegel, E. (2013) *The Power of Predictive Analytics*. Hoboken, NJ: John Wiley & Sons.
- Simpson, L. (2002) Great Expectations, *Training*, 39(2) 40–44.
- Sims, R. R. (1993) Evaluating public sector training programs, *Public Personnel Management*, 22(4), 591–615.
- Singh, V., Vinnicombe, S. & Kumra, S. (2006) Women in formal corporate networks: an organisational citizenship perspective, *Women in Management Review*, 21(6), 458–482.
- Smith, A. (1990) Evaluation of management training subjectivity and the individual, *Journal of European Industrial Training*, 14(1), 12–15.

- Smith, A. J. & Piper, J. A. (1990) The tailor-made training maze: a practitioner's guide to evaluation, *Journal of European Industrial Training*, 14(8), 2–24.
- Smithson, M. & Verkuilen, J. (2006) *Fuzzy Set Theory*. Thousand Oaks, CA: Sage Publications.
- Spicer, D. P. & Sadler-Smith, E. (2006) Organisational learning in smaller manufacturing firms, *International Small Business Journal*, 24(2), 133–158.
- Spitzer, D. R. (1984), 'Why Training Fails', *Performance Instruction Journal*, 9, 6–11.
- Spitzer D. R. 2005. Learning effectiveness measurement: a new approach for measuring and managing learning to achieve business results. *Advanced Development of Human Resources*. 7:55–70
- Sridharan, S., Campbell, B. & Zinzow, H. (2006) Developing a stakeholder-driven anticipated timeline of impact for evaluation of social programs, *American Journal of Evaluation*, 27, 148–162.
- Stake, R. E. (1994) *The Art of Case Study Research*, Thousand Oaks, CA: Sage.
- Steele, K and Stefánsson, H. (2015). *Decision Theory*, The Stanford Encyclopedia of Philosophy (Winter 2015 Edition).
- Steiger, J. H. & Lind, J. M. (1980) Statistically based tests for the number of common factors: *Annual Meeting of the Psychometric Society*, 28 May, Iowa City, IA, Psychometric Society, Iowa City, IA.
- Stevens, J. P. (1992) *Applied Multivariate Statistics for the Social Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Stewart, T. A. (1997) *Intellectual Capital*. New York, NY: Doubleday/Currency.
- Stokking, K. (1998) Levels of evaluation: Kirkpatrick, Kaufman and Keller, and beyond, *Human Resource Development Quarterly*, 7(2), 179–183.
- Storey, J. (1992) *Management of Human Resources*. Oxford: Blackwell.
- Strauss, A. & Corbin, J. M. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage.
- Strauss, A. (1997) *Qualitative Analysis for Social Scientists*. Cambridge: Cambridge University Press.

- Strebel, P. & Keys, T. (2005) *Mastering Executive Education: How to Combine Content with Context and Emotion, the IMD Guide*. Harlow: Pearson Education.
- Stufflebeam, D. L. (1983) The CIPP model for program evaluation, in Madaus, G.F., Scriven, M. S. & Stufflebeam, D. L. (eds), *Evaluation Models*. Boston, MA: Kluwer-Nijhoff Publishing.
- Sturman, M. C. (2000) Implications of utility analysis adjustments for estimates of human resource intervention value, *Journal of Management*, 26(2), 281–299.
- Surowiecki, J. (2004) *The Wisdom of Crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*, 1st edn. New York, NY: Doubleday.
- Sutton, B. (2006). Adopting a holistic approach to the valuation of learning programmes deployed in corporate environments. DProf Thesis: Middlesex University Library
- Swanson, R. A. & Holton, E.F., III. (1999) *Results: How to Assess Performance, Learning, and Perceptions in Organisations*. San Francisco, CA: Berrett-Koehler Publishers.
- Swanson, Richard A. & Sleezer (1987) Training Effectiveness Evaluation, *Journal of European Industrial Training*, 11(4), 7–16.
- Swenson, G. (1992) Fuzzy Logic: A Tutorial & Trends, *Futurics: A Quarterly Journal of Futures Research*, 16(3/4), 45.h
- Tabachnick, B. G. & Fidell, L. S. (2007) *Using Multivariate Statistics*, 5th edn. Boston, MA: Pearson Education.
- Tabassi, A. A., Mahyuddin, R. & Abu, H. A. B. (2012) Effects of Training and Motivation Practices on Teamwork Improvement and Task Efficiency: The Case of Construction Firms, *International Journal of Project Management*, 30(2), 213–224.
- Tai, W. (2006) Effects of Training Framing, General Self-Efficacy and Training Motivation on Trainees' Training Effectiveness, *Personnel Review*, 35(1), 51–65.
- Talbot, C. (1992) Evaluation and validation: a mixed approach, *Journal of European Industrial Training*, 16(5), 26–32.
- Talbot, M. (1991) *The Holographic Universe*. New York, NY: HarperCollins.
- Taleb, N. N. (2012) *Antifragile: Things That Gain from Disorder*. New York, NY: Random House.

- Tamkin, P., Yarnall, J. & Kerry, M. (2002), *Kirkpatrick and Beyond: A Review of Models of Training Evaluation, Report 392*. Brighton: Institute of Employment Studies.
- Tannenbaum, S. I. & Yukl, G. (1992), Training and development in work organisations, *Annual Review of Psychology*, 43, 399–441.
- Tannenbaum, S. I., Mathieu, J. E., Salas, E. & Cannon-Bowers, J. A. (1991) Meeting trainee's expectations: The influence of training fulfilment on the development of commitment, self-efficacy and motivation, *Journal of Applied Psychology*, 76, 759–769.
- Tasca, J. E., Ensslin, L., Ensslin, S. R. & Martins Alves, M. B., (2010), An approach for selecting a theoretical framework for the evaluation of training programs, *Journal of European Industrial Training*, 34(7), 631–655.
- Taylor, F. W. (1911) *The Principles of Scientific Management*. New York, NY: Harper & Brothers.
- Tenbrunsel, A. E., Wade-Benzoni, K. A., Messick, D. M. & Bazerman, M. H. (2000) Understanding the Influence of Environmental Standards on Judgments and Choices, *The Academy of Management Journal*, 43(5), 854–866.
- Tennant, C., Boonkrong, M. & Roberts, P. (2002) The design of a Training Programme Measurement Model, *Journal of European Industrial Training*, 26(5), 230–240.
- Tesoro, F. (1998) Implementing an ROI measurement process at Dell Computer, *Performance Improvement Quarterly*, 11, 103–114.
- Thach, E. C. (2002) The impact of executive coaching and 360° feedback on leadership effectiveness, *Leadership and Organisation Development Journal*, 23, 205–221.
- Thompson, J. & Cole, M. (1999) Strategic competency — The learning challenge, *Journal of Workplace Learning*, 7(5), 153–162.
- Toffler, A., et al. (1972) *The Futurists*. New York, NY: Random House.
- Torres, R. T., Preskill, H. S. & Piontek, M. (1996) *Evaluation Strategies for Communicating and Reporting: Enhancing Learning in Organizations*. Thousand Oaks, CA: Sage.
- Towler, A. J. & Dipboye, R. L. (2009) Effects of Trainer Expressiveness, Organisation and Training Orientation on Training Outcomes, *Journal of Applied Psychology*, 86(4), 664–673.

- Tracey, B. J. & Michael, J. (1995) Training Effectiveness, *Cornell Hotel and Restaurant Administration Quarterly*, 36(6), 36–43.
- Tracey, J. B. Tannenbaum, S. I. & Kavanagh, M. J. (1995) Applying trained skills on the job: the importance of the work environment, *Journal of Applied Psychology*, 80, 239–252.
- Training Magazine. (2015) *2015 Training Industry Report*, <https://trainingmag.com/trgmag-article/2015-training-industry-report> [retrieved on 12th Jan 2016].
- Tsai, W. C. & Tai, W. (2003) Perceived Importance as a mediator of the relationship between Training Assignment and Training Motivation, *Personnel Review*, 32(2), 151–163.
- Tzeng, G., Chiang, C. & Li, C. (2007) Evaluating intertwined effects in e-learning programs: a novel hybrid MCDM model based on factor analysis and DEMATEL, *Expert Systems with Applications*, 32, 1028–1044.
- U.S. Department of Health, Education and Welfare. (1975) *A practical guide to measuring project impact on student achievement*. Washington, DC: U.S. Government Printing Office.
- Ulrich, D. (1997) *Human Resource Champions*. Boston, MA: Harvard Business Press.
- Van der Sluis, L., Williams, R. & Hoeksema, L. (2002) Measuring the quality of managerial learning on the job, *Human Resource Management*, 13(8), 1266–1278.
- Van Velsor, E. & Ascalon, E. (2008) The role and impact of leadership development in supporting ethical action in organisations, *Journal of Management Development*, 27(2), 187–195.
- Vance, R. J. & Colella, A. (1990) Effects of two types of feedback on goal acceptance and personal goals, *Journal of Applied Psychology*, 75(1), 68–76.
- Vassen, J. (2006) Programme theory evaluation: multicriteria decision aid and stakeholder values, *Evaluation*, 12, 397–417.
- Velda, A., Caetano, A., Michel, J. W., Lyons, B. D. & Kavanagh, M. J. (2007) The effects of training design, individual characteristics and work environment of the transfer of learning, *International Journal of Training and Development*, 11(4), 284–294.

- Wade, J. B., O'Reilly, C. A. & Pollock, T. G. (2006) Overpaid CEOs and Underpaid Managers: Fairness and Executive Compensation, *Organization Science*, 17(5), 527–544.
- Wallace, T. L. (2008) Integrating participatory elements into an effectiveness evaluation, *Studies in Educational Evaluation*, 34(4), 201–207.
- Walonick, D. S. (1993). *Survival Statistics*. Northfield, MN: StatPac Inc.
- Walster, E., Walster, G. W. & Berscheid, E. (1978) *Equity Theory and Research*. Boston, MA: Allyn and Bacon.
- Wang, G. G. & Spitzer, D. R. (2005) Human resources development measurement and evaluation: looking back and moving forward, *Advances in Developing Human Resources*, 5(5), 1–15.
- Warr, P., Bird, M. & Rackham, N. (1970) *Evaluation of Management Training: A Practical Framework, with Cases, for Evaluating Training Needs and Results*. London: Gower Press.
- Warr, P., Bird, M. & Rackham, N. (1978) *Evaluation of management training*. London: Glover.
- Wason, P. C. (1960) On the failure to eliminate hypotheses in a conceptual task, *Quarterly Journal of Experimental Psychology*, 12, 129–140.
- Waterman, H. (1998) Embracing ambiguities and valuing ourselves: issues of validity in action research, *Journal of Advanced Nursing*, 28(1), 101–105.
- Watkins, R., Leigh, D., Foshay, R. & Kaufman, R. (1998) Kirkpatrick plus: evaluation and continuous improvement with a community focus, *ETR&D*, 46(4), 90–96.
- Weinrich, P. (2001) *Decision Space: Multidimensional Utility Analysis*. Cambridge: Cambridge University Press.
- Weiss, C. H. (1998) *Evaluation: Methods for Studying Programs and Policies*. Upper Saddle River, NJ: Prentice Hall.
- Weiss, C. H. (1983) The stakeholder approach to evaluation: origins and promise, *New Directions for Program Evaluation*, 17, 3–14.
- Weissner, C. A., Hatcher, T., Chapman, D. & Storberg-Walker, J. (2008) Creating new learning at professional conferences: an innovative approach to conference learning, knowledge construction and evaluation, *Human Resource Development International*, 11(4), 367–383.

- Weisstein, E. W. (n.d.) Strong Law of Large Numbers, *MathWorld*, <http://mathworld.wolfram.com/StrongLawofLargeNumbers.html> [retrieved on 12th Jan 2014].
- West, K. D. (2006) Forecast Evaluation, in Granger, C., Elliott, G. & Timmermann, A. (eds), *Handbook of Economic Forecasting*, Vol. 1. Philadelphia, PA: Elsevier.
- Western, S. (2008) *Leadership: A Critical Text*. London: Sage.
- White, J. P. (1982) *The Aims of Education Restated*. London: Routledge and Kegan Paul.
- Whyte, G. & Latham, G. (1997) The futility of utility analysis revisited — When even an expert fails, *Personnel Psychology*, 50(3), 601–610.
- Wiley, J. W. & Brooks, S. M. (2000) The high-performance organisational climate: How workers describe top-performing units, in Ashkanasy, N. M., Wilderom, C. P. M. & Peterson, M.F. (eds), *Handbook of Organizational Culture & Climate*, 177–191. Thousand Oaks, CA: Sage Publications.
- Williams, R. (1969) Can We Evaluate Management Training?, *Industrial Training International*, 4(3), 120–122.
- Wills, G. & Oliver, C. (1996), Measuring the ROI from management action learning, *Management Development Review*, 9(1), 17–21.
- Wills, S. (1993) Evaluation concerns: a systematic response, *Journal of European Industrial Training*, 17(10), 10–14.
- Wolters, C. A. (1998) Self-regulated learning and college students' regulation of motivation, *Journal of Educational Psychology*, 90, 224–235.
- Wood, R., Bandura, A. & Bailey, T. (1990) Mechanisms governing organisational performance in complex decision-making environments, *Organizational Behavior and Human Decision Processes*, 46(2), 181–201.
- Wright, P. M., George, J. M., Farnsworth, S. R. & McMahan, G. C. (1993) Productivity and extra-role behavior: The effects of goals and incentives on spontaneous helping, *Journal of Applied Psychology*, 78(3), 374–381.
- Wu, W. (2008) Dimensions of social capital and firm competitiveness improvement: the mediating role of information sharing, *Journal of Management Studies*, 45(1), 122–146.

- Ya Hui Lien, B., Yu Yuan Hung, R. & McLean, G. N. (2007) Training evaluation based on cases of Taiwanese benchmarked high-tech companies, *International Journal of Training and Development*, 11(1), 35–48.
- Yeo, R. K. (2008) How does learning (not) take place in problem-based learning activities in workplace contexts?, *Human Resource Development International*, 11(3), 317–330.
- Yin, R. K. (2003) *Case Study Research: Design and Methods*, 3rd edn. Thousand Oaks, CA: Sage.
- Zaciewski, R. D. (2001) Measuring Training's Effectiveness, *Quality Progress*, 34(6), 36–42.
- Zajonc, R. B. (1980) Feeling and Thinking: Preferences need no Inferences, *American Psychologist*, 35(2), 151–175.